National Library of Canada
Bibliothèque nationale du Canada

Canadian Theses Service
Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

NL-339 (r. 88/04) c

Canada

UNIVERSITY OF ALBERTA

PROBING MOLECULAR INTERACTIONS AT VARIOUS LEVELS OF STRUCTURAL DEFINITION

by

NATALIE C.J. STRYNADKA

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOCHEMISTRY

EDMONTON, ALBERTA

FALL 1990

National Library of Canada
Bibliothèque nationale du Canada

Canadian Theses Service
Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-64794-9

605 Third Avenue
New York, NY 10158-0012
212.850.6000
FAX 212.850.6088
Telex 12.7063
Cable JONWILE

John Wiley & Sons, Inc.
Publishers Since 1807

October 12, 1990

Natalie C.J. Strynadka
University of Alberta
Department of Biochemistry
474 Medical Sciences Building
Canada T6G 2H7

Dear Ms. Strynadka:

Thank you for your letter dated October 1, 1990, requesting permission to reuse "Two Trifluoperazine-Binding Sites on Calmodulin Predicted from Comparative Molecular Modelling with Troponin-C" from PROTEINS: STRUCTURE, FUNCTION & GENETICS 3:1-17 (1988) and "Model for the Interaction of Amphiphilic Helices with Troponin-C and Calmodulin " from PROTEINS: STRUCTURE, FUNCTION & GENETICS 7:234-248 (1990) to appear in a doctoral thesis.

Please be aware that although you assigned rights to this work to us in your contributor agreement, you retained the following rights:

1. The right to make copies for your own teaching use.
2. The right to reuse the articles or parts thereof in a collection of your own works or in a textbook that you are writing or in lecture notes, press releases, and reviews.
3. The right to reuse figures and tables in other articles that you may submit for publication as journal articles, review articles, or chapters in books, with the proviso that full bibliographic references be given to this particular article.

We would appreciate you crediting Wiley-Liss, a division of John Wiley & Sons, Inc. The following components must be included: Title, author(s) and/or editor(s), journal title (if applicable), Copyright (year and owner).

If you have any questions please call (212)850-6014.

Yours truly,

Kim Harvey

Kim Harvey
Permissions Assistant

Directors
William R. Brian
Sidney D. Drell
Sandra M. Faber
Eugene Garfield
James F. Hewett
William Kaufmann
Daniel E. Koshland, Jr.
Joshua Lederberg
Gardner Lindzey
J. Murray Luck (Emeritus)
William F. Miller
Charles Yanofsky)
Richard N. Zare
Harriet A. Zuckerman

ANNUAL REVIEWS INC.
A NONPROFIT SCIENTIFIC PUBLISHER

4139 El Camino Way
P.O. Box 10139
Palo Alto, California 94303-0897 USA
Telephone 415-493-4400
Fax 415-855-9815

President and Secretary-Treasurer
John S. McNeil

Editor-In-Chief
William Kaufmann

Publication Manager
Mickey G. Hamilton

Business Manager
Donald S. Svedeman

Publication Manager
Ann B. McGuire

October 1, 1990

FAX COVER SHEET

RECIPIENT'S FAX #: 408-492-0886

TO: Dr. Natalie C.J. Strynadka
Dept. of Biochemistry
University of Alberta

FROM: Mrs. Judith A. Mueller JM
Copyrights and Permissions
ANNUAL REVIEWS INC.

NUMBER OF PAGES (including cover sheet): 1 *

MESSAGE: Dear Dr. Strynadka: This is to confirm that you have permission to use ARI copyrighted material, specifically the chapter entitled "Crystal Structures of the Helix-Loop-Helix Calcium-Binding Proteins," which appeared in the Annual Review of Biochemistry, Volume 58, c 1989, as a chapter in your thesis, with the proviso that you use the following acknowledgement on the first page of the reprinted material:

> "Reproduced, with permission, from the Annual Review of Biochemistry, Volume 58, c 1989, by Annual Reviews Inc."

Thank you for obtaining permission from Annual Reviews Inc.

*IF YOU DID NOT RECEIVE ALL OF THE PAGES TRANSMITTED, PLEASE CALL THE FAX OPERATOR AS SOON AS POSSIBLE. THANK YOU.

ANNUAL REVIEWS OF:
Anthropology
Astronomy and Astrophysics
Biochemistry
Biophysics and Biophysical Chemistry
Cell Biology
Computer Science
Earth and Planetary Sciences
Ecology and Systematics
Energy
Entomology
Fluid Mechanics
Genetics
Immunology
Materials Science
Medicine
Microbiology
Neuroscience
Nuclear and Particle Science
Nutrition
Pharmacology and Toxicology
Physical Chemistry
Physiology
Phytopathology
Plant Physiology and Plant Molecular Biology
Psychology
Public Health
Sociology

University of Alberta
Edmonton

Department of Biochemistry

Canada T6G 2H7

474 Medical Sciences Building, Telephone (403) 492-2422
Fax (403) 492-0886

October 1, 1990

RECEIVED
OCT 02 1990
JOURNALS DIVISION

Journals Division
Journal of Virology
ASM
1325 Massachusetts Avenue, N.W.
Washington, D.C. 20005-4171
U.S.A.

Attention: Linda Illig

Dear Dr. Illig:

I am writing to gain copyright permission of a paper entitled "The Use of Synthetic Peptides to Map the Antigenic Determinants of Glycoprotein D of Herpes Simplex Virus" (Strynadka *et al.*, (1988), J. Virol. 62: 3474-3483) for use in my doctoral thesis. Unfortunately, I have only just been informed that these letters of copyright permission must be submitted to the University Registrar before October 10, 1990 in order that I may convocate in the fall session. As I am quite desperate to convocate at this time, I would greatly appreciate any effort your agency would undertake in sending a letter of copyright permission for the above mentioned paper before October 10. I feel express mail will be required. If it is not too inconvenient, could you please send the permission letter by courier or express mail C.O.D. to our departmental address.

Thank you for your time and consideration in this matter.

Sincerely,

Natalie C.J. Strynadka

NCJS/dl

PERMISSION GRANTED CONTINGENT ON AUTHOR PERMISSION
AND APPROPRIATE CREDIT
American Society for Microbiology
Journals Division
L. M. Illig Date 10-2-90

UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR:	Natalie C.J. Strynadka
TITLE OF THESIS:	Probing Molecular Interactions at Various Levels of Structural Definition
DEGREE:	Doctor of Philosophy
YEAR THIS DEGREE GRANTED:	Fall 1990

Permission is hereby granted to the UNIVERSITY OF ALBERTA LIBRARY to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

Signature Natalie S.T.

Permanent Address

4 GARNET AVE

Sherwood Park, Alta

Date: Oct 12 / 1990

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled "**Probing Molecular Interactions at Various Levels of Structural Definition**" submitted by Natalie C.J. Strynadka in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Michael James

Michael N.G. James, F.R.S.

D G Scraba

Douglas G. Scraba

Lawrence B. Smillie

Lawrence B. Smillie

David L Tyrrell

David L. Tyrrell

David C Phillips

Sir David Phillips, K.B.E., F.R.S.

Date: SEPT 26, 1990

— For Jay and Brian —

ABSTRACT

The predictive algorithm Surfaceplot [Parker, Guo & Hodges, *Biochemistry* 25: 5425-5432 (1986)] was used to examine glycoprotein D of herpes simplex virus, type 1 (HSV-1) for amino acid residues with a high probability of being exposed on the molecular surface. Based upon these data, 11 different peptides corresponding to 10-residue segments in the primary sequence of glycoprotein D, and one 20-residue segmentm, were synthesized, conjugated to carrier proteins and used to generate specific antisera in rabbits. Two synthetic peptides predicted not to be on the surface of glycoprotein D were included as "negative controls". The polyclonal antisera against individual synthetic peptide conjugates were in turn evaluated for their ability to recognize both isolated glycoprotein D and intact HSV-1 virions in ELISA. Eight linear antigenic sites on glycoprotein D were thereby defined out of the twelve antipeptide antisera prepared based upon Surfaceplot predictions; four of these contain epitopes to which complement-independent neutralizing antibodies can be generated. The latter sites correspond to sequences 12-21, 267-276, 288-297 and 314-323 of the mature protein. An additional peptide sequence, 2-21, was found to generate antisera which had potent virus neutralizing capacity in the presence of complement. Identification of a neutralizing epitope in the sequence 314-323 makes it likely that the membrane spanning region of glycoprotein D is within the subsequent sequence 323-339.

Among the known regulatory proteins that are conformationally sensitive to the binding of calcium ions, calmodulin and troponin-C have the greatest primary sequence homology. This observation has led to the conclusion that the most accurate predicted molecular model of calmodulin would be based on the X-ray crystallographic coordinates of the highly refined structure of turkey skeletal troponin-C. This paper describes the structure of calmodulin built from such a premise. The resulting molecular model was subjected to conjugate gradient energy minimization to remove unacceptable intramolecular non-bonded contacts. In the analysis of the resulting structure, many features of calmodulin, including the detailed conformation of the Ca^{2+}-binding loops, the amino- and

carboxy-terminal hydrophobic patches of the Ca^{2+}-bound form and the several clusters of acidic residues can be reconciled with much of the previously published solution data. Calmodulin is missing the N-terminal helix characteristic of troponin-C. The deletion of three residues from the central helical linker (denoted D/E in troponin-C) shortens the molecule and changes the orientation of the two domains of calmodulin by 60° relative to those in troponin-C. The molecular model has been used to derive two possible binding sites for the antipsychotic drug trifluoperazine, a potent competitive inhibitor of calmodulin activity.

Crystals of troponin C are stabilized by an intermolecular interaction that involves the packing of helix A from the N-terminal domain of one molecule onto the exposed hydrophobic cleft of the C-terminal domain of a symmetry related molecule. Analysis of this molecular recognition interaction in troponin C suggests a possible mode for the binding of amphiphilic helical molecules to troponin C and to calmodulin. From the template provided by this troponin C packing, it has been possible to build a model of the contact region of mastoporan as it might be bound to the two Ca^{2+} binding proteins. A possible binding mode of melittin to calmodulin is also proposed. Although some of the characteristics of binding are similar for the two amphiphilic peptides, the increased length of melittin requires a significant bend in the central helix similar to that suggested recently for the MLCK calmodulin binding peptide [Persechini & Kretsinger, *J. Cardiovascular. Pharm.* **12**:501-512 (1988)]. Not only are the hydrophobic interactions important in this model, there are several favorable electrostatic interactions that are predicted as a result of the molecular modelling. The regions of troponin C and calmodulin to which amphiphilic helices bind are similar to the regions to which the neuroleptic drugs such as trifluoperazine have been predicted to bind [Strynadka & James, *Proteins: Structure, Function and Genetics* **3**:1-17 (1988)].

A structure of the trisaccharide 2-acetamido-2-deoxy-D-muramic acid-β(1→4)-2-acetamido-2-deoxy-D-glucose-β(1→4)-2-acetamido-2-deoxy-D-muramic acid (NAM-

NAG-NAM), bound to subsites B, C and D in the active site cleft of hen egg white lysozyme has been determined and refined at 1.5 Å resolution. The resulting atomic coordinates indicate that the NAM residue at site D is distorted from the full $^4C_1$ chair conformation to one in which the ring atoms C1, C2, O5 and C5 are approximately coplanar, and the hydroxymethyl group is positioned axially (a conformation best described as a sofa). This finding supports the original proposals that suggested the ground state conformation of the sugar bound at site D is strained to one which more closely resembles the geometry required for the oxocarbonium-ion transition state, the next step along the reaction pathway [Phillips, D.C., *Scientific American* **215**:78-90 (1966)]. Additionally, detailed analysis at 1.5 Å resolution of the environments of the catalytic residues, Glu35 and Asp52, lends added information on the properties which may allow lysozyme to promote the stabilization of an unusually long-lived oxocarbonium-ion transition state.

Intermolecular interactions between the N-acetyl-muramic acid residue in site D and the lysozyme molecule that contribute to the saccharide ring distortion include: close packing of the O3' lactyl group with a hydrogen-bonded "platform" of enzyme residues (Asp52, Asn46, Asn59, Ser50 and Asp48), a close contact between the hydroxymethyl group of ring D and the 2'-acetamido group of ring C and a strong hydrogen-bonded interaction between the NH of Val109 and O6 of ring D which stabilizes the observed quasi-axial orientation of the $-CH_2OH$ group. Additionally, the structure of this complex shows a strong hydrogen bond between the carboxyl group of Glu35 and the β-anomeric hydroxyl group of the NAM residue in site D. The hydrogen-bonded environment of Asp52 in the native enzyme and in the complex coupled with the very unfavorable direction of approach of the potential carboxylate nucleophile makes it most unlikely that there is a covalent glycosyl-enzyme intermediate on the hydrolysis pathway of hen-egg white lysozyme.

ACKNOWLEDGEMENTS

It is not possible for me to fully express my gratitude to Mike James and Anita Sielecki. The time in their lab has truly revolutionized the way I see the world, scientifically and otherwise. Mike's incredible insight on all aspects of protein structure and function coupled with Anita's careful, honest approach to the processing and analysis of crystallographic data create a scientific standard in the lab from which one can't help to learn from and aspire to. They have been teacher, referee, counsellor and friend.

Other members of the "James Gang" have also played an important role in my time as a graduate student. Marie Fraser has been a kind friend, fellow skiier, and a source of knowledge on many aspects of crystallography and computing. Koto Hayakawa patiently introduced me to the techniques of protein crystallization and precession camera work. Stan Moore shared many of the trials of my first attempts at Area Detector Data Collection. His humorous way of always looking at the lighter side of things made these and other difficult times in the lab much easier. In addition to Anita, Mark Israel, Dave Bacon, Colin Broughton, Masao Fujinaga and Randy Read have provided a legacy of computer programs which made the job of structural analysis on a day-to-day basis an increasingly easier task. John Moult kindly introduced me to the world of energy minimization and molecular dynamics. Collectively, the "Gang" has provided a warm, stimulating environment within which to work and play.

Mae Wylie has been an essential member in the production of this thesis and my published papers. She dealt with the many revisions of manuscripts expertly and patiently and with a speed that always managed to outwit any deadline.

Through the years I have come to value the informal, collaborative nature of the University of Alberta Department of Biochemistry. I have learned greatly from the techniques and advice given freely by members of the laboratories of Dr. Bob Hodges, Dr. Doug Scraba, Dr. Larry Smillie, and Dr. Cyril Kay. This type of collaborative spirit has also taught me the value of looking at a scientific problem from different perspectives.

The friendship of fellow students has helped me through the ups and downs of graduate work. Jennifer van Eyck, Iris Leung, and Pam Werner shared everything from strange results to mountain hikes.

Finally, many thanks to my family — George, Ellen, and Todd Strynadka, who have always encouraged and supported me along my chosen path.

This research has been made possible by the Medical Research Council of Canada through a grant to the MRC Group in Protein Structure and Function. I have been supported as a graduate student by an Alberta Heritage Foundation for Medical Research Studentship.

TABLE OF CONTENTS

Page

CHAPTER 1: INTRODUCTION 1

REFERENCES 3

CHAPTER 2: THE USE OF SYNTHETIC PEPTIDES TO MAP THE ANTIGENIC DETERMINANTS OF GLYCOPROTEIN D OF HERPES SIMPLEX VIRUS 4

INTRODUCTION 4

MATERIALS AND METHODS 5

Surface Profile Prediction 5

Synthesis of Peptides 6

Conjugation to Carrier Proteins 8

Production of Antisera 8

Cells and Virus 9

Purification of Glycoprotein D 9

Enzyme-linked Immunosorbent Assays (ELISA) 10

Neutralization Tests 11

Immunofluorescence 12

RESULTS 12

Selection of Potential Linear Antigenic Sites 12

Affect of Linkers and Coupling Position on Antisera Production 14

Antibody Titer 14

Antibody Reactivity with HSV-1 Virions 15

Neutralization Tests 15

Reactivity of Human Sera with Synthetic Peptide Antigens 16

HSV Type-Specific Immune Response 17

DISCUSSION 18

ACKNOWLEDGEMENTS 20

REFERENCES 21

CHAPTER 3: CRYSTAL STRUCTURES OF THE HELIX-LOOP-HELIX CALCIUM-BINDING PROTEINS 33

INTRODUCTION 34

FUNCTIONAL OVERVIEW 36

Troponin C 36

Calmodulin 37

Parvalbumin 38

Intestinal Calcium Binding Protein 39

STRUCTURAL OVERVIEW 39

Primary Structures 39

Tertiary Structure 41

CRYSTALLOGRAPHIC ANALYSES 41

OVERALL ARCHITECTURE 42

Troponin C 44

Calmodulin 44

Parvalbumin 46

Intestinal Calcium Binding Protein 46

THE CA^{2+} COORDINATION 47

Comparison of Loop Conformations 49

The Standard Ca^{2+}-Binding Loop 50

N-Terminal Loops of TnC 56

The Variant Loop of ICaBP 57

HELICES 58

Helix Regularity 58

Helix Packing 62

HYDROPHOBIC INTERACTIONS 63

ELECTROSTATIC INTERACTIONS 66

CONFORMATIONAL CHANGE OF TNC 68

ACKNOWLEDGEMENTS 69

REFERENCES 71

CHAPTER 4: TWO TRIFLUOPERAZINE BINDING SITES ON CALMODULIN PREDICTED FROM COMPARATIVE MOLECULAR MODELLING WITH TROPONIN-C 103

INTRODUCTION 103

METHODS 105

RESULTS AND DISCUSSION 106

Caveats Concerning this Predicted Model of CaM 107

Description of the General Structure of CaM 109

(A) CA^{2+}-SATURATED FORM 109

(B) CA^{2+}-FREE FORM 109

Description of the Ca^{2+} Binding Loops 110

Factors Influencing the Ca^{2+} Binding Affinity 111

Description of the Hydrophobic Patches 113

Model of Trifluoperazine Binding to CaM 114

CONCLUSIONS 118

REFERENCES 120

CHAPTER 5: A MODEL FOR THE INTERACTION OF AMPHIPHILIC HELICES WITH TROPONIN C AND CALMODULIN 135

INTRODUCTION 135

METHODS 137

RESULTS AND DISCUSSION 138

Analysis of the Crystal Packing of Troponin C 138

Analysis of the X-ray Crystallographically Defined Interaction of the A-

Helix with the C-Terminal Domain of Troponin C 138

Analysis of the Modelled Interaction of Mastoporan with the C-Terminal Domain of Troponin C 140

Analysis of the Modelled Interaction of Mastoporan with the C-Terminal Domain of Calmodulin 143

Binding of Amphiphilic Helices to the N-Terminal Domain of CaM 144

A Model for the Interaction of Melittin Binding to CaM 147

CONCLUSION 150

ACKNOWLEDGEMENTS 151

REFERENCES 152

CHAPTER 6: LYSOZYME REVISITED: CRYSTALLOGRAPHIC EVIDENCE FOR DISTORTION OF AN N-ACETYL-MURAMIC ACID RESIDUE BOUND IN SITE D 168

ABSTRACT 169

INTRODUCTION 170

MATERIALS AND METHODS 172

Crystallization and Data Collection 172

Refinement of Native Lysozyme 174

Structure Solution and Refinement of the HEWL-MGM Complex 174

RESULTS 175

Quality of the Refined Structures 175

Comparison of HEWL in the Native and Complexed Form 178

PROTEIN ATOMS 178

SOLVENT MOLECULES 179

Structural Features of the HEWL-MGM Complex 180

SITE B N-ACETYLMURAMIC ACID 180

SITE C N-ACETYLGLUCOSAMINE 181

SITE D N-ACETYL MURAMIC ACID 182

Comparison of the Refined HEWL-MGM with Previous Complexes of HEWL 185

Environments of the Catalytic Residues 187

GLUTAMIC ACID 35 187

ASPARTIC ACID 52 188

Implications of the HEWL-MGM Complex for the Catalytic Mechanism of Lysozyme 189

DISTORTION IN SITE D 189

IS THERE COVALENT GLYCOSYL-ENZYME INTERMEDIATE? 193

ACKNOWLEDGEMENTS 195

REFERENCES 196

LIST OF TABLES

		Page
Table 2.1	Surfaceplot predicted surface sites and antigenic sites predicted using Hopp and Woods (1981) parameters	24
Table 2.2	Sequences of HSV-1 glycoprotein D peptides synthesized	25
Table 2.3	Sequences of peptides synthesized to investigate the effect of linker position and type of linker on the immune response to synthetic peptide conjugates	26
Table 2.4	Results of antibody binding to HSV-1 virions and neutralization assays	27
Table 2.5	Sequences of synthetic peptides based on type-specific regions of glycoprotein D of herpes simplex virus	28
Table 3.1	Calcium binding proteins crystal structures	81
Table 3.2	Primary sequences	82
Table 3.3	Summary of crystallographic data	84
Table 3.4	Ca^{2+}-binding domain comparisons	86
Table 3.5	Superpositions of pairs of loops	87
Table 3.6	Interhelix angles and contact distances	88
Table 5.1	Alignment of TnC/CaM inhibitory peptides with the A-helix of TnC	157
Table 5.2	Summary of the interactions of the observed and modelled amphilic helices with TnC	158
Table 6.1	Summary of observed intensities for the lysozyme-MGM complex	201
Table 6.2	Refinement statistics	202
Table 6.3	Average temperature factors for saccharides bound to sites B, C and D of HEWL	203
Table 6.4	Geometry of glycosidic linkages	204
Table 6.5	Summary of hydrogen bonded contacts in the HEWL-MGM complex	205

LIST OF FIGURES

		Page
Figure 2.1	Surfaceplot of HSV-1 glycoprotein D	29
Figure 2.2	ELISA reactivity of peptide antisera with HSV-virions	30
Figure 2.3	Immunofluorescence of peptide antisera with HSV virions	31
Figure 2.4	Glycoprotein D secondary structure prediction	32
Figure 3.1	Schematic of a HLH Ca^{2+} binding domain	89
Figure 3.2	Main chain representation of HLH proteins	90
Figure 3.3	Pentagonal bipyramidal calcium coordination	92
Figure 3.4	Calcium binding loops of the HLH proteins	93
Figure 3.5	Hydrogen bonding in HLH calcium binding loops	95
Figure 3.6	Helices in HLH proteins	96
Figure 3.7	Hydrophobic domains in HLH proteins	98
Figure 3.8	Schematic of the Ca^{2+} induced conformational change	101
Figure 3.9	The N-terminal domain of TnC in the Ca^{2+} free and Ca^{2+} filled state	102
Figure 4.1	Amino acid sequence alignment of HLH proteins	105
Figure 4.2	Backbone representation of TnC and CaM	126
Figure 4.3	All atom representation of CaM	127
Figure 4.4	Backbone representation of apo CaM	127
Figure 4.5	Calcium binding loops in CaM	128
Figure 4.6	The N-terminal domain of CaM in the apo and Ca^{2+} filled form	129
Figure 4.7	C-terminal domain of CaM in the apo and Ca^{2+} filled form	130
Figure 4.8	Chemical configuration of TFP	131
Figure 4.9	TFP binding to the N-terminal domain of CaM	132
Figure 4.10	TFP binding to the C-terminal domain of CaM	133
Figure 4.11	Binding of two TFP molecules to CaM	134
Figure 5.1	Ribbon diagram of CaM	159

Figure 5.2	Space filling diagram of TnC	160
Figure 5.3	Hydrophobic domains of TnC and CaM	162
Figure 5.4	Interaction of symmetry related TnC molecules	164
Figure 5.5	Interaction of amphiphilic helices with TnC and CaM	165
Figure 5.6	Interaction of mellitin with CaM	167
Figure 6.1	Schematic of a hexasaccharide substrate	206
Figure 6.2	Overall fold of lysozyme	206
Figure 6.3	Proposed mechanism of lysozyme	207
Figure 6.4	Ramachandran plot of lysozyme	209
Figure 6.5	Temperature factor plot of lysozyme	209
Figure 6.6	Electron density of the lysozyme active site	210
Figure 6.7	Electron density of MGM	211
Figure 6.8	Overlap of lysozyme in native and MGM bound forms	212
Figure 6.9	Space filling representation of lysozyme in the native and MGM bound forms	213
Figure 6.10	Hydrogen bonding interactions between protein and sugar	214
Figure 6.11	Schematic of a NAM sugar conformation	215
Figure 6.12	Difference density of the site D NAM	216
Figure 6.13	Overlap of NAM in the isolated and protein bound form	216
Figure 6.14	Comparison of MGM and other saccharides	217
Figure 6.15	Environment of glutamic acid 35	218
Figure 6.16	Environment of aspartic acid 52	219

CHAPTER 1

INTRODUCTION

I have been fortunate during the course of my graduate work to learn and utilize a number of biochemical and biophysical techniques which probe protein function at varying levels of structural definition. Three protein systems have been investigated: glycoprotein D of herpes simplex virus type I; the calcium-modulated proteins troponin C and calmodulin; and the antibacterial enzyme lysozyme. In the first system, the viral glycoprotein D, the only structural information available was limited to that of a primary sequence. Clusters of amino acids within this sequence were analyzed to predict possible regions of the protein that could lie on the molecular surface and thus be likely candidates as antigenic determinants. Peptides mimicing the predicted regions were synthesized and conjugated to appropriate carriers. Antipeptide antibodies were raised and analyzed for their ability to bind to glycoprotein D in the isolated and virion bound form. In addition, the antibodies were tested for their ability to neutralize viral infectivity with the aim of testing the effectiveness of the synthesized peptides as potential viral vaccines.

In the second system studied, the known three dimensional structures of troponin-C (Herzberg & James, 1988) and calmodulin (Babu *et al.*, 1988) were used as the basis for computer-aided modelling of the molecular complexes of these proteins with their target molecules. The wealth of existing biochemical information on the calcium dependent nature of these molecular interactions was of invaluable assistance in the modelling process. These studies allowed us to predict possible binding modes of inhibitory drugs and substrates as well as conformational changes that may be required for complexation to occur.

Finally, the third system, hen egg white lysozyme provided a high resolution, refined crystallographic structure of both a native protein and the complex with one of its natural substrates. In light of my own modelling work on the calcium binding proteins described in Chapters 4 and 5, it was specially gratifying to find that the detailed

information provided by these 3-dimensional crystal structures verified several of the predictions that were advanced on the basis of much earlier modelling studies (Phillips, 1966; Blake *et al.*, 1967). In addition, it provided a valuable lesson on the amount and type of additional information one can glean from analyses carried out at almost atomic resolution. These three fields of study are obviously distinct. For this reason more detailed introductory prefaces of the individual projects are given in each of the following chapters.

REFERENCES

Babu, Y.S., Bugg, C.E. & Cook, W.J. 1988. Structure of calmodulin refined at 2.2 Å resolution. *J. Mol. Biol.* **204**:191-204.

Blake, C.C.F., Johnson, L.N., Mair, G.A., North, A.C.T., Phillips, D.C. & Sarma, V.R. 1967. *Proc. Roy. Soc. London B* **167**, 378-388.

Herzberg, O. & James, M.N.G. 1988. Refined crystal structure of troponin C from turkey skeletal muscle at 2.0 Å resolution. *J. Mol. Biol.* **203**:761-779.

Phillips, D.C. 1966. *Scientific American* **215**:78-90.

CHAPTER 2

THE USE OF SYNTHETIC PEPTIDES TO MAP THE ANTIGENIC DETERMINANTS OF GLYCOPROTEIN D OF HERPES SIMPLEX VIRUS[1]

INTRODUCTION

There has been increasing experimental support in recent years for the concept of using synthetic peptides as viral vaccines. The basic premise is that peptides representing surface regions of a viral coat protein can give rise to antibodies of selected specificity which can confer humoral immunity to that virus. For example, antigenic peptides have been synthesized which correspond to protein segments of influenza A (Muller *et al.*, 1982), polio (Emini *et al.*, 1983) and hepatitis B (Neurath *et al.*, 1985) viruses. In the case of foot-and-mouth disease virus, synthetic peptides corresponding to amino acid residues 151-160 and/or 200-213 of the capsid protein VP1 have been shown to elicit the production of virus neutralizing antibodies in cattle (Bittle *et al.*, 1982; DiMarchi *et al.*, 1986).

Virus-neutralizing antibodies have also been generated in mice and rabbits using peptide-conjugates of the N-terminal region of glycoprotein D of herpes simplex virus (Eisenberg *et al.*, 1985). Herpes simplex virus type 1 (HSV-1) is known to cause a number of human diseases, including facial lesions, corneal keratitis and encephalitis (Nahmias *et al.*, 1981). Glycoprotein D of HSV-1 is the major protein component of the envelope of the mature virion and plays an important role in initiating the infectious process (Spear, 1985). It is therefore a good candidate for the synthetic approach to immunization.

Parker *et al.* (1986) have developed a predictive algorithm, named Surfaceplot, which can be used to scan a protein amino acid sequence for residues which should be exposed on the surface of the molecule. Application of Surfaceplot to glycoprotein D of HSV-1 revealed 15 linear segments (5 or more residues) of the polypeptide which have a

---

[1] A version of this chapter has been published. Strynadka *et al.*, 1988. *J. Virol.* 62:3474-3483. Relevant material published subsequent to the paper is given in Chapter 7.

high probability of surface exposure. Such segments are potentially antigenic (Benjamin *et al.*, 1984). Surfaceplot also identified a number of segments of glycoprotein D with a very low probability of surface exposure; two of these were selected for control experiments. Fourteen peptides corresponding to the predicted surface and interior segments were synthesized, photochemically linked to protein carriers and injected into rabbits. The polyclonal sera produced in response to these peptides were then screened by an enzyme-linked immunosorbent assay (ELISA) for reactivity against peptide-conjugates, purified glycoprotein D and HSV-1 virions.

The results communicated in this paper show that all of the peptides selected produced positive immune responses (ELISA) in rabbits, and that all the rabbit antisera reacted with isolated glycoprotein-D. Eight of the antisera reacted with intact herpes virions, confirming the presence of these sequences on the surface of glycoprotein D *in situ*. In plaque-reduction (neutralization) tests, five of the antisera inhibited virus infection. The greatest virus neutralization capacity was associated with antisera raised to the peptide-conjugate of the N-terminal amino acid residues 2-21. Strong neutralization was also produced by antisera to the sequence 267-276.

MATERIALS AND METHODS

Surface Profile Prediction

Synthetic peptides were chosen from surface regions predicted from the algorithm of Parker *et al.* (1986). This algorithm (Surfaceplot) combines the HPLC hydrophilicity parameters of Guo *et al.* (1986), the accessibility parameters of Janin (1979) and the flexibility parameters of Karplus and Schultz (1985). The HPLC parameters were derived from the retention times in reversed-phase chromatography of model synthetic peptides, Ac-Gly-X-X-$(Leu)_3$-$(Lys)_2$-amide, where X was substituted with each of the twenty amino acids found in proteins. The accessibility parameters derived from X-ray crystallographic data from twenty-two proteins represent the partition coefficient of each amino acid residue between the surface and the inside volume of the protein. The flexibility parameters were

determined from the X-ray crystallographic data of the average temperature values of all C_α-atoms of each amino acid in thirty-one proteins. It has been shown that a combination of these three parameters improved the prediction of surface regions compared to using any single parameter alone (Parker *et al.*, 1986).

Synthesis of Peptides

All chemicals and solvents used were reagent grade. Co-poly(styrene 2% divinylbenzene) benzhydryl amine resin (hydrochloride salt, substitution = 0.8 mmole/gm of resin) was purchased from Protein Research Foundation, Osaka, Japan. Dichloromethane was distilled from calcium carbonate. Trifluoroacetic acid was purchased from Halocarbon, Hackensack, N.J., p-benzoyl-benzoic acid from Aldrich Chem. Co., Milwaukee, WI., and N^α-t-Boc-L-lysine from Sigma Chemical Co.,St. Louis, MO. All other protected amino acids were purchased from Institute Armand Frappier, Laval, Quebec.

Peptides were synthesized on a Beckman peptide synthesizer Model 990 using the general procedure outlined by Erickson and Merrifield (1976) for solid-phase synthesis, with the modifications described by Parker and Hodges (1985). Amino groups were protected at the α-amino position with a t-butyloxycarbonyl (Boc) group, and the following side chain blocking groups were used: Lys(2-chlorobenzoxycarbonyl), Thr (benzyl), Glu (O-benzyl), Asp(O-benzyl), Arg (tosyl), Cys (methoxybenzyl) and Tyr (2,6-dichlorobenzyl). A radioactive label Boc 1-^{14}C glycine, prepared as described previously (Worobec *et al.*, 1983), was introduced at an appropriate glycine in the sequence. Since the peptide corresponding to residues 12-21 of glycoprotein D was to be used to investigate different peptide-protein conjugation procedures other than the photochemical coupling method described below, two analogs were synthesized, one with a cystine added to the N-terminus and a second with a tyrosine added to the N-terminus.

The photochemical coupling reagents N^ε-(p-benzoylbenzoyl)-N^α-t-butyloxycarbonyl-lysine, Boc-Lys(BB) (Parker and Hodges, 1985), and p-benzoylbenzoic acid,

BBA, were coupled to the N-terminal α-amino group of the peptide-resin or Boc-Lys(BB) was coupled as the first C-terminal amino acid during synthesis. To obviate the limited solubility of these reagents in dichloromethane the following procedure was found to give high reproducible coupling yields. Boc-Lys(BB) or BBA (1600 μmoles/800 μmole peptide resin) was dissolved in dimethylformamide (200 - 500 μl) diluted with dichloromethane (4.5 ml) and cooled to 4°C. Dicyclohexylcarbodiimide (DCC, 800 μmoles) was added and the reaction was left at 4°C for 30 min before being filtered and added to the deprotected and neutralized peptide resin. After stirring for 30 min, an additional portion of DCC (800 μmoles) indichloromethane (1 ml) was added and the reaction was stirred for one hour. All peptides were synthesized as C-terminal amides. When Lys(BB) was synthesized at the C-terminus, the N-terminal residue of the completed peptide was deprotected, neutralized and acetylated with acetic anhydride/pyridine/toluene (1:2:2,v:v:v).

Peptides were cleaved from the resin support using hydrofluoric acid (HF):10% anisole (v:v) at 4° for 45 min (Hodges and Merrifield, 1975a). The solvents were removed under reduced pressure at 4°, the resin was washed with ether and the peptide was extracted with TFA (4 x 5 ml). The combined TFA washes were evaporated, and the peptide redissolved in water and lyophilized.

Purification of peptides was accomplished by reversed-phase high performance liquid chromatography (HPLC) using a Whatman Partisil CCS/C8 column (Whatman, Clifton, NJ, 250 x 4.6 mm I.D.) or a Synchropak RP-P C18 reversed-phase column (250 x 10 mm I.D.). A gradient of 1% B/min was used at a flow rate of either 1 ml/min (C8 column) or 2 ml/min (C18 column). Solvent A was 0.1% trifluoroacetic acid in water and solvent B was 0.05% trifluoroacetic acid in acetonitrile. The absorbance was recorded at two wavelengths, 210 nm for the peptide bond and 260-270 nm for the photoprobe (BB or BBA).

Amino acid analysis were performed on a Durrum Model D-500 high pressure

automatic analyzer after hydrolysis of samples with 6 N HCl:0.1%phenol in sealed, evacuated tubes at 110° for 24 h. When peptide-resin hydrolysates were required the peptide-resins were hydrolyzed as described by Hodges and Merrifield, 1975b. The mean of the molar ratios of all accurately measurable amino acids in the acid hydrolysate was used to calculate the concentration of the peptide.

Conjugation to Carrier Proteins

Concentrated solutions of carrier proteins were prepared in the following manner. To 2 g of bovine serum albumin (BSA) (Fraction V; Sigma) was added 1 ml of 0.1 M phosphate buffer, (pH 6.8), and the gummy mixture allowed to stand overnight before being further diluted to 4 ml (1 mg/2 µl). Similarly, 500 mg of keyhole limpet hemocyanin (KLH) (Sigma) was mixed with 1 ml of 0.1 M $NaHCO_3$ and allowed to stand overnight before being further diluted to 5 ml (1 mg/10 µl).

Photochemical coupling of peptide to carrier was accomplished by dissolving or suspending lyophilized peptide (1 to 2 mg) in a minimum volume of phosphate buffer (10-30 µl of 0.1 M KH_2PO_4, pH 6.8) and adding ~ 10 mg of BSA or KLH from the stock solutions, (20-100 µl). The photolysis was carried out for 45 min in a Rayonet RPR 208 preparative apparatus (Southern New England Ultraviolet Co., Middletown, CT) equipped with 350 nm lamps. The reactor was placed in a cold room and a constant air temperature (8-10°) surrounding the sample was maintained by an electric fan mounted in the bottom of the reactor. Following photolysis the conjugates were dissolved in 4 ml 8 M urea, dialyzed (Spectrapor tubing, 6000-8000 MW cutoff) against 0.1 M NH_4HCO_3 (once) and 0.025 M NH_4HCO_3 (twice), and lyophilized. Peptide conjugation to proteins via the peptide N-terminal cysteine or tyrosine was carried out according to the methods of Liu *et al.* (1979) and Bassiri (1980), respectively. Peptide:protein ratios were ascertained from the specific activity of the radioactively-labeled peptide, or by amino acid analysis, and ranged from 5 to 15 moles of peptide per mole of carrier protein.

Production of Antisera

Antisera to synthetic peptide-KLH conjugates were obtained from 6 week old male New Zealand white rabbits. On day 1, 40 µg of peptide-KLH conjugate (in sterile 0.9% NaCl, 500 µl) was thoroughly mixed with an equal volume of Freund's complete adjuvent (Difco Laboratories, Detroit, MI.) and injected (200 µl per injection) into the left and right subscapular regions and gluteal muscle. The same procedure was repeated at day 14 except that 200 µg of peptide (in sterile 0.9% NaCl, 500 µl) in incomplete Freund's adjuvant (500 µl) was used as before (200 µl per injection). The rabbits were bled from the ear vein at day 28 and reinnoculated as described for day 14. This bleeding/inoculation protocol was carried out for a total of 6 weeks.

Cells and Virus

Human embryonic lung cells (Flow 2000 line; Flow Laboratories, McLean VA.) were cultured in 175 cm^3 plastic flasks (Corning Glass, Corning, N.Y.), containing Eagle's basal medium (BME; Flow Labs) supplemented with 10% (v:v) fetal calf serum (Gibco, Grand Island, N.Y.). Herpes simplex virus type 1 (strain Kos) was used to infect confluent monolayers at an input multiplicity of infection of between 1 and 5. Infection was allowed to proceed for 18 hr. The infected cells were scraped from the plastic surface, resuspended in phosphate buffered saline (PBS) (Dulbecco & Vogt, 1954) and homogenized (Dounce homogenizer, close-fitting pestle). Following centrifugation (400 *g*, 10 min), the pelleted material was resuspended, re-homogenized and recentrifuged (1800 *g*, 10 min). The two supernatants were combined and centrifuged at 100,000 g for 1 hr (Beckman Ti60 rotor). The virion-containing pellet was resuspended in a minimal volume of PBS, layered onto a discontinuous gradient of sucrose (20%, 40%, 60% w/w) and centrifuged (Beckman SW 27.1 rotor) at 100,000 *g* for 1 hr. The opalescent virion band (at ~ 45% sucrose) was removed from the side of the tube with a hypodermic syringe. The presence of intact virions was confirmed by electron microscopy.

Purification of Glycoprotein D

N-octyl glucopyranoside (Sigma) was added to the supernatants obtained from homogenized HSV-1 infected cells to a final concentration of 1% (w:v). Potentially infectious DNA was inactivated by incubating with 1 mg DNase I/ml (Sigma) for 30 min at 37°. This lysate was then dialyzed overnight (4°) against PBS to remove the detergent. An affinity column of Concanavalin A-Sepharose (Pharmacia, Dorval, Que.) was prepared and equilibrated with 50 mM Tris-HCl, (pH 7.2), containing 0.5 mM $MgCl_2$ and 250 mM $MnCl_2$. The dialyzed HSV-1 lysate was then applied to the column washed with 10 volumes of buffer and the bound glycoproteins subsequently eluted with 0.2 M mannose (Aldrich). The eluant was dialyzed overnight at 4°C against water, then lyophilized. The resulting powder was dissolved in 1 ml of water and subjected to isocratic HPLC on a Superose 12 gel filtration column (molecular weight range 10^3 - 10^5 daltons; Pharmacia) using 0.1 M phosphate buffer (pH 7.0) 50 mM NaCl as eluting buffer. Protein elution was monitored by absorption at 280 nm, and glycoprotein D was found to be eluted as a single peak after approximately 30 min. The purity of this product was confirmed by SDS-PAGE. Samples of the 30 min peak were electrophoresed on 10% polyacrylamide gels (Laemmli, 1970). The gels were fixed in methanol:acetic acid (40:10; v/v) and stained for protein using a protein-silver stain kit (Biorad Laboratories, Richmond, CA.). Two bands of apparent molecular masses 60 kD and 52 kD were detected and correspond to the apparent molecular masses reported for mature glycoprotein D (59-65 kD) and pre-glycoprotein D (52 kD) (Spear, 1976; Eisenberg *et al.*, 1979).

Enzyme-linked Immunosorbent Assays (ELISA)

Antisera titres were determined using a modification of the ELISA protocol described by Voller *et al.* (1974). Microtitre plates (Immunolon II; Dynatech, Alexandria, VA) were coated with 100 µl of peptide-BSA conjugate (50 µg/ml)in 0.5 M Na_2CO_3 (pH 9.6) for 16 hrs at 4°C in a moist chamber. This represented a saturating concentration of protein carrier. Peptide-BSA conjugates were used in the ELISA wells when KLH

peptide-conjugates were used for immunization in order to measure antipeptide antibody titre and to exclude anticarrier antibody titre. The antigen coated wells were washed (3 x) with PBS containing 0.05% (v:v) Tween 20 (PBS-T). After washing, the plates were incubated for 2 hr at room temperature with 100 µl/well of the appropriate dilution of antisera in PBS-T containing 0.5% (w:v) BSA. Unbound or excess Ab was removed by washing with PBS-T (3 x) and the plate was incubated with goat anti-rabbit IgG conjugated to alkaline phosphatase (1:1000 dilution; Boehringer-Mannheim Corp., Dorval, Quebec). After washing with PBS-T (3 x), p-nitrophenyl phosphate (Sigma 104 substrate) in 10% (v:v) diethanolamine was added. Thirty to sixty min later the absorbance at 410 nm was measured on a Titerek Multiscan plate reader (Flow Laboratories, Mississauga, Ontario). Titres from ELISA assays expressed as the negative logarithm of the reciprocal of the highest dilution of serum which still gave a positive response. In ELISA assays for glycoprotein D and virion reactivity, the same protocol was followed except for coating the plates with 100 µl (10 µg/ml) of glycoprotein D or HSV-1 preparations. Acute and immune sera from individuals with HSV-1 infections were generously provided by Mr. R. Devine of the Alberta Provincial Laboratory, Edmonton, Alberta.

Neutralization Tests

Prior to analysis in plaque reduction assays, the immunoglobulin component of each antiserum was separated from complement and other serum proteins by affinity chromatography on DEAE-Affigel Blue columns (Biorad). This was done in order to estimate the contribution that complement makes to virus neutralization under our experimental conditions. Dilutions of antibody with and without the addition of 5% guinea-pig complement, were incubated for 1 hr at room temperature with 100 and 200 plaque-forming units of virus. These mixtures were then assayed for unneutralized virus using the plaque technique of Wentworth and French (1969). Briefly, monolayers of fibroblast cells were prepared by seeding 0.75 x 10^6 cells in each well of 6-well tissue culture plates (Costar; Cambridge, MA). After overnight incubation, the growth medium

was removed and each plate inoculated with 0.4 ml of the neutralization mixture. Inocula were allowed to adsorb for 30 min at room temperature after which time 4 ml of overlay medium (MEM containing 0.1% $NaHCO_3$ and 0.3% agar) was added. When the overlay had solidified, the plates were incubated at 37° in a 5% CO_2 atmosphere for 72 hrs. The monolayers were fixed using 4% neutral buffered formalin prior to the removal of the overlay, and subsequently stained with 1% crystal violet. Clear plaques of dead cells were enumerated.

Immunofluorescence

Confluent monolayers of Flow 2000 cells on glass coverslips were infected with HSV-1 or HSV-2 at an input multiplicity of between 1 and 5 pfu/cell. Culturing was for 24 hr after infection, at which time the cells were washed twice with PBS and fixed in methanol. After drying, the coverslips were stored at -20° until assayed. The coverslips were then incubated for 1 hr at room temperature with 10% normal goat serum in PBS to block HSV-induced Fc receptors, washed with PBS, and covered with a 1:500 dilution of the peptide-specific polyclonal rabbit sera. After 1 hr the coverslips were washed (3 x) with PBS and then incubated for 1 hr with fluorescein isothiocyanate-labeled (FITC) goat anti-rabbit IgG (1:32 dilution; Sigma). Following a subsequent wash the cells were mounted in a mixture of 1 part PBS and 9 parts glycerol. Negative controls of pre-immune rabbit sera were included for each test, and photography was carried out with a Leitz Ortholux II fluorescent microscope using Kodacolor VR 400 film.

RESULTS

Selection of Potential Linear Antigenic Sites

It is now generally accepted that the entire surface of a protein is antigenic (Benjamin *et al.*, 1984). We have developed a surface profile algorithm, Surfaceplot, and have shown for several proteins that there is a good correlation of its predictions with surface sites determined from immunological and X-ray structural data (Parker *et al.*,

1986). Surfaceplot has been applied to glycoprotein D of HSV-1, using the amino acid sequence reported by Watson *et al.* (1982). The computer-generated surface profile of glycoprotein D is shown in Figure 2.1B and the predicted surface residues are listed in Table 2.1. A predicted large buried region (amino acid residues 314-339) corresponds to the proposed membrane-spanning area, and two large surface areas (340-351 and 356-365) correspond to the proposed hydrophilic C-terminal cytoplasmic tail. In the external N-terminal region, the Surfaceplot predicts five major (sequences of 10 or more residues) linear surface regions (12-21, 83-93, 130-140, 244-275 and 299-313) and eight minor (sequences of 5 or more residues) linear surface regions (28-33, 44-51, 73-78, 116-120, 144-148, 184-189, 206-213, 224-229) for a total of 13 predicted surface regions of 5 or more residues (denoted by an asterisk or dagger in Table 2.1). By comparison, analysis using Hopp and Woods (1981) hydrophilicity parameters (see Fig. 2.1A) reveals only four antigenic sites in the external region of glycoprotein D of 5 or more residues (denoted by an asterisk in Table 2.1). As shown in Table 2.1, the sequences predicted using Hopp and Woods parameters exist as smaller subsets of those predicted using Surfaceplot. We have synthesized peptides corresponding to all major predicted surface sequences which were 10 or more residues in length (Table 2.2, denoted by a dagger). Peptides 52-61 and 150-159 were synthesized as negative controls since these regions were predicted not to be on the surface (dashed arrows, Fig. 2.1B). Due to the large number of predicted minor surface sites or sites containing less than 5 surface residues (Table 2.1, 24 out of 31) we selected a representative sampling of these sites throughout the sequence for synthesis as described below. The peptides (181-190, 206-215) were chosen since they contained minor predicted surface sequences which were 5 or more residues in length (Table 2.2, denoted by an asterisk). Peptides (63-72, 121-130, 314-323) contained one or two single predicted surface residues and peptide 288-297 contained a surface site of four residues (Table 2.2; solid arrows in Fig. 2.1B). Sequence 314-323 was chosen since it contained only one predicted surface residue (site 27, Table 2.1) and is within the twenty-five amino acid

sequence (314-339) that was predicted by Watson *et al.* (1982) to be the transmembrane region. The peptides synthesized (Table 2.2) containing the two minor or five major surface sites are shaded in Fig. 2.1B. It should be noted that of the 13 surface sites represented by 12 synthetic peptides (Table 2.2) from the prediction of Surfaceplot, seven of these sites were not predicted by the program of Hopp and Woods (Table 2.1).

Effect of Linkers and Coupling Position on Antisera Production

Five methods of conjugation of peptide to carrier protein were investigated (Table 2.3). All five of these conjugates elicited a strong antibody response in rabbits as tested in ELISA against peptide-conjugate antigen. In the controls, there was no ELISA reaction when the plates were coated with the coupling reagents BB-Lys or BB-Gly. Peptide-conjugates with photoprobe attached to the ε-amino group of the N-terminal lysine residue (BB-Lys 12-21) or photoprobe attached to the ε-amino group of the C-terminal lysine residue (12-21 BB-Lys) showed significantly reduced reactivity in ELISA assays with isolated glycoprotein D or within tact virions, in comparison with the reactivity produced by a peptide-conjugate with the photoprobe attached to the N-terminal α-amino group of glycine (BB-Gly 12-21). Since the method of coupling, the photoprobe and the sequence of peptide are identical, these results suggest that the position of the photoprobe provides a different anti-peptide antibody response even though the titres against peptide-conjugates were similar. The chemically linked peptides, Cys 12-21 and Tyr 12-21, also showed a positive immune response in ELISA assays against isolated glycoprotein D and intact virions. The titres produced by peptides chemically linked to carriers were similar to those of the photoprobe linked peptide, BB-Gly 12-21. Given these results and the ease of coupling with the photochemical reagent, all subsequent experiments were done with BB-Gly peptide-conjugates.

Antibody Titre

The ELISA titres (expressed as the negative logarithm of the reciprocal of the

highest dilution of serum which gave a positive response) of antisera raised against BB-Gly peptide-KLH conjugates and tested against BB-Gly peptide-BSA conjugates were ~ 4-5. All antisera were also found to react with isolated glycoprotein D, with a mean titre of 3. The positive reactivity of control peptide antisera (52-61, 150-159) with isolated glycoprotein D (ELISA titers of <1.7; *i.e.* 1:50 dilution) indicated that isolated glycoprotein D was no longer in its native conformation.

Antibody Reactivity with HSV-1 Virions

Interaction of the peptide antisera with purified virions is shown in Figure 2.2. Eight of the peptide antisera reacted with titres > 2 (*i.e.* 1:100 dilution) with this antigen. It is interesting that antibodies raised to the peptides which showed low titres to the whole virion did react well with isolated glycoprotein D. This particular result indicates that either a conformational change altering antigenicity occurs during incorporation of glycoprotein D into the virion envelope, or that close packing masks an antigenic site. Antisera raised to the sequence 314-323 also shows reactivity with the whole virion. It has been proposed (Watson, 1982) that this sequence is likely to be part of the transmembrane region. The reactivity of the antisera raised to peptide 314-323 (which was predicted by Surfaceplot to have surface residues) would place the start of the transmembrane region closer to the C-terminus. Antisera raised to the control sequence (52-61, 150-159) did not react with the intact HSV-1 virion.

Neutralization Tests

Antisera which reacted with intact virions in ELISA was screened for virus-neutralization capacity using a plaque-reduction test. Antibodies to five of the eight surface regions were found to elicit a neutralizing response (Table 2.4). The strongest neutralization was associated with the antisera to N-terminal sequence 2-21 in the presence of guinea pig complement. Without complement the neutralization associated with these antisera was much weaker. A fragment of sequence 2-21, namely residues 12-21,

neutralized virus equally with or without complement. Since the extent of this neutralization was equal to that of antisera 2-21 in the absence of complement, one may conclude that sequence 2-21 contains two epitopes; one in the sequence 12-21 and the other, which requires complement for neutralization, elsewhere within residues 2-21.

The rabbit antisera raised to predicted surface regions 267-276 and 288-297 also contained neutralizing antibodies. It is interesting to note that in the plaque reduction assay, the antisera to 267-276 exhibited the strongest complement-independent response.

It was surprising to find neutralization associated with the antisera to peptide 314-323. As we have stated previously, our results suggest that this sequence lies adjacent to the transmembrane region of glycoprotein D.

Reactivity of Human Sera with Synthetic Peptide Antigens

The recognition of the synthetic peptide antigens by sera obtained from (11) individuals with acute or convalescent (3 and 8 individuals, respectively) HSV-1 infections was assessed using ELISA assays. For acute sera, although the titre against viral antigen was found to be significant (ELISA titer >2), no strong response to any one peptide was observed. It is known that in the primary stages of herpesvirus infection, it is primarily the cytotoxic T cells and cell-mediated immune responses (Nash *et al.*, 1985) which combat infection. The lack of development of humoral response to linear antigenic determinants was not, therefore, unexpected.

The response of immune sera in some cases showed a lack of specificity for surface sites; for example, three sera reacted with the predicted internal sequences 52-61 and two reacted with 150-159. This reactivity was probably produced by a fortuitous cross reaction of non-HSV human antibodies with these sequences. The N-terminal sequence (2-21) was recognized (ELISA titer >2) by 6 of the 8 sera. The sequence 267-276 was recognized by all 8 convalescent sera. This may be indicative of the immunological importance of these sites, and these are potential candidates for a synthetic vaccine. De Freitas *et al.* (1985) have reported that a synthetic peptide corresponding to residues 8-23 activated human

peripheral blood T-cells. A vaccine containing this sequence and the sequence 267-276 might be expected to elicit both a strong humoral immune response and to prime cell-mediated immunity to herpes simplex virus.

HSV Type-Specific Immune Response

The interaction of antisera raised to type-specific sequences with cells infected with HSV-1 or HSV-2 was detected by indirect fluorometric assay. Three type-specific sequences from HSV-1 (those in which there are a number of non-conservative amino-acid changes between corresponding peptides for glycoprotein D of HSV-1 and HSV-2) have been synthesized (sequences 38-47, 83-92 and 199-208; Table 2.5). These type-specific peptides also contained residues predicted to be at or near the protein surface. Antisera to these sequences showed bright cytoplasmic fluorescence when assayed against HSV-1 infected cells (Figure 2.3). Although these sequences were found to have little surface exposure on the intact HSV virion, as judged by a low ELISA titre with virions (Table 2.5) they did react with denatured glycoprotein D. This would suggest that the fluorescence that was found in the cytoplasm of type 1 infected cells, was due to the binding of antibody to glycoprotein D that was as yet unincorporated into the virion structure. In fluorometric analysis of cells infected with HSV-2 (strain 333), antisera to peptide 83-92 produced a pattern of fluorescence identical to that obtained with HSV-1 infected cells; *i.e.* bright pools of cytoplasmic fluorescence. Antiserum to peptide 38-47 produced plasma membrane fluorescence, similar to that which we have previously observed with non-specific or cross-reactive antibody binding to membrane components. The antiserum to peptide 199-208 was non-reactive. Controls of antisera raised to the type-common peptide sequences showed identical patterns of fluorescence with both types of herpes simplex virus infection. Negative controls of pre-immunization rabbit sera gave low overall fluorescence with small centers of apparently non-specific binding. The immunofluorescence results suggest that the amino acid changes in peptide 83-92 do not significantly affect antibody binding, whereas the differences in peptides 38-47 and 199-208 profoundly affect antibody binding.

The crossreactivity of antisera in the 83-92 region is somewhat surprising; however, the major differences occurring in the center of the sequence are homologous (E-D-V for D-E-A). Antiserum to peptide 199-208 could conceivably be useful as a type-specific diagnostic reagent, permitting one to differentiate between HSV-1 and HSV-2 infections.

DISCUSSION

We used the Surfaceplot predictive algorithm to select linear amino acid sequences in HSV-1 glycoprotein D which have a high probability of exterior location and are, therefore, potentially antigenic. Ten-residue peptides corresponding to eleven such sequences, a twenty residue peptide containing two surface sites, and two "negative controls" (*i.e.* sequences with very low probability of being on the surface of glycoprotein D) were synthesized, conjugated to KLH carrier protein and used to immunize rabbits. The polyclonal antisera thereby obtained was found, in each case, to react positively in ELISA with the peptide conjugated to BSA and with isolated glycoprotein D. These results demonstrated that all the synthetic peptide-conjugates were immunogenic in rabbits and that glycoprotein D isolated by affinity chromatography and HPLC and bound to the ELISA plate is in a denatured state. In contrast, not all of the antisera reacted positively in ELISA with HSV virions; that is, with glycoprotein D *in situ*. Eight linear antigenic sites on HSV-1 virions were identified. These correspond to glycoprotein D sequences 12-21, 2-21, 121-130, 206-215, 267-276, 288-297, 303-312 and 314-323.

In the Surfaceplot predictions, sequences 121-130 and 314-323 had less than 5 contiguous surface residues (Table 2.1). If one examines a Chou-Fasman secondary structure representation (Chou and Fasman, 1974a, b; 1978) of glycoprotein D (Figure 2.4), the sequence 121-130 includes a glycosylation site (Asn 121) and parts of two β-turns. These criteria are obviously indicative of a surface segment, and this demonstrates that such features as glycosylation sites and β-turns should be considered along with Surfaceplot data in selecting potential antigenic regions. In the case of peptide sequence 314-323, Surfaceplot indicates positive values for residues 322 and 324 (Table

2.1), and the Chou-Fasman structure (Fig. 2.4) predicts β-turns for residues in the regions 312-315 and 323-324. The fact that at least part of the sequence 314-323, being antigenic, must find itself on the exterior of the viral envelope makes it necessary to revise the proposal by Watson *et al.* (1982) that the membrane-spanning region for glycoprotein D is residues 314-339 (or residues 340-364 of the glycoprotein D precursor protein). Since Eisenberg *et al.* (1985) have described an epitope within the sequence 340-356 in isolated glycoprotein D, we suggest that the membrane spanning region is most likely to follow the predicted β-turn at residues 323-324 and extend to approximately residue 340. There are no charged residues within this sequence; part of it (residues 326-330) may be helical and the remainder is probably random coil, even though Chou-Fasman predictions suggest β-sheet.

Synthetic peptides corresponding to sequences 2-21, 12-21, 267-276, 288-297 and 314-323 elicited a virus-neutralizing immune response in rabbits. The long peptide encompassing residues 2-21 was found to comprise two epitopes. One of these lies within the sequence 12-21, which by itself elicits antibodies capable of neutralizing virions in the absence of complement. Peptide 2-21 retains the ability to generate complement-independent neutralizing antibody, and in addition elicits the production of rabbit antibodies which neutralize virus very efficiently in the presence of complement. These results are compatible with those obtained by Cohen *et al.* (1984) who demonstrated that rabbits immunized with synthetic peptide 8-23 developed HSV-1 neutralizing antibodies. Eisenberg *et al.* (1985) reported that antiserum to peptide 268-287 had no neutralizing activity; in contrast we found that antisera to peptide 267-276 showed the strongest complement-independent neutralizing activity of all of the five linear epitopes which we identified.

Dietzschold *et al.* (1984) raised polyclonal and monoclonal antibodies to glycoprotein D isolated from HSV-1 and HSV-2. Interaction of these antibodies with synthetic peptides (peptide 1-23 with citraconylated lysines 10 and 20, and peptides 8-23 and 11-23) suggested that one of the epitopes in the N-terminal 1-23 sequence is located

between residues 11 and 19. We confirmed this suggestion since antiserum to peptide 12-21 reacts with isolated glycoprotein D and with HSV-1 virions.

Eisenberg *et al.* (1985) identified another antigenic site in the sequence 268-287 by showing reactivity of this peptide with an anti-glycoprotein D monoclonal antibody. This monoclonal antibody did neutralize HSV-1 at a dilution of 1:50; however, when peptide 268-287 was conjugated to KLH and used to immunize rabbits, the antipeptide antibodies did not react with either glycoprotein D or intact virions. We found, in contrast, that antibodies to the KLH conjugate of peptide 267-276 recognized both glycoprotein D and HSV-1 virions. They also neutralized. This difference might be explained on the basis that the longer peptide (268-287) adopts a conformation distinct from that of the native epitope, whereas the shorter peptide (267-276) is more flexible. An alternative explanation is that the method of conjugating the peptide to KLH resulted in the differences in its immunogenicity. Eisenberg *et al.* (1985) coupled their peptide via a C-terminal cysteine, whereas we coupled our peptide through an N-terminal BB-gly. These results point out the importance of selection of cross-linker and point of attachment of peptide to carrier protein in the design of peptide vaccines.

ACKNOWLEDGEMENTS

This work was supported by grants from the Medical Research Council to R.S.H. and D.G.S. N.C.J.S. and M.J.R. were recipients of studentship awards from the Alberta Heritage Foundation for Medical Research.

We wish to thank Beverly Bellamy and Perry d'Obrenan for assisting with the preparation of the manuscript and Roger Bradley for his excellent technical assistance.

REFERENCES

BASSIRI, R.M., DVORAK, J., & UTIGER, R.D. (1979). Method of radioimmunoassay. In Methods of Hormone Immunoassay, pp. 46-47. Edited by B.M. Jaffe & H.R. Behrman. Academic Press, New York.

BENJAMIN, D.C., BERZOFSKY, J.A., EAST, I.J., GURD, F.R.N., HANNUM, C., LEACH, S.J., MARGOLIASCH, E., MICHAEL, J.G., MILLER, A., PRAGER, E.M., REICHLIO, M., SERCATZ, E.E., SMITH-GILL, S.J., TODD, P.E. & WILSON, A.C. (1984). The antigenic structure of proteins: A reappraisal. *Annu. Rev. Biochem.* **2**, 67-101.

BITTLE, J.L., HOUGHTEN, R.A., ALEXANDER, H., SHINNICK, T.M., SUTCLIFFE, J.G., LERNER, R.A., ROWLANDS, D.J. & BROWN, F. (1982). Protection against foot-and-mouth disease by immunization with a chemically synthesized peptide predicted from the viral nucleotide sequence. *Nature, (London)* **298**, 30-34.

CHOU, P.Y. & FASMAN, G.D. (1974a). Conformational parameters of amino acids in helical β-sheet and random coil regions. *Biochem. (U.S.A.)* **13**, 211-222.

CHOU, P.Y. & FASMAN, G.D. (1974b). Prediction of protein conformation. *Biochem. (U.S.A.)* **13**, 222-245.

CHOU, P.Y. & FASMAN, G.D. (1978). Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol.* **47**, 45-148.

COHEN, G.H., DIETZCHOLD, B., PONCE DE LEON, M., LONG, D., GOLUB, E., VARRICHIO, A. PEREIRA, L. & EISENBERG, R.J. (1984). Localization and synthesis of an antigenic determinant of herpes simplex virus glycoprotein D that stimulates the production of neutralizing antibody. *J. Virol.* **49**, 102-108.

DE FREITAS, E.C., DIETZSCHOLD, B. & KOPROWSKI, H. (1985). Human T-lymphocyte response *in vitro* to synthetic peptides of herpes simplex virus glycoprotein D. *Proc. Natl. Acad. Sci. USA* **82**, 3425-3429.

DIETZSCHOLD, B., EISENBERG, R.J., PONCE DE LEON, M., GOLUB, E., HUDECZ, F., VARRICHIO, A. & COHEN, G.H. (1984). Fine structure analysis of type-specific and type-common antigenic sites of herpes simplex virus glycoprotein D. *J. Virol.* **52**, 431-435.

DIMARCHI, R., BROOKE, G., GALE, C., CRACKNELL, V., DOEL, T. & MOWAT, N. (1986). Protection of cattle against foot-and-mouth disease by a synthetic peptide. *Science* **232**, 639-641.

DULBECCO, R. & VOGT, M. (1954). Plaque formation and isolation of pure lines with poliomyelitis viruses. *J. Exp. Med.* **99**, 167-182.

EISENBERG, R.J., HYDREAN-STERN, C. & COHEN, G.H. (1979). Structural analysis of precursors and products forms of type-common envelope-glycoprotein D (CP-1 antigen) of herpes simplex virus type 1. *J. Virol.* **31**, 608-620.

EISENBERG, R.J., LONG, D., PONCE DE LEON, M., MATTHEWS, J.T., SPEAR, P.G., GIBSON, M.G., LASKY, L.A., BERMAN, P., GOLUB, E. & COHEN, G.H. (1985). Localization of epitopes of herpes simplex virus type 1 glycoprotein D. *J. Virol.* **53**, 634-644.

EMINI, E.A., JAMESON, B.A. & WIMMER, E. (1983). Priming for and induction of anti-poliovirus neutralising antibodies by synthetic peptides. *Nature (London)* **304**, 699-703.

ERICKSON, B.W., & MERRIFIELD, R.B. (1976). Solid-phase peptide synthesis. In The Proteins, vol. II, pp. 647-666. Edited by H. Nevrath and R.L. Hill. New York:Academic Press.

GUO, D., MANT, C.T., TANEJA, A.K., PARKER, J.M.R. & HODGES, R.S. (1986). Prediction of peptide retention times in reversed-phase HPLC: Determination of retention coefficients of amino acid residues of model synthetic peptides. *J. Chromatography* **359**, 499-517.

HODGES, R.S. & MERRIFIELD, R.B. (1975a). The role of serine-123 in the activity and specificity of ribonuclease. *J. Biol. Chem.* **250**, 1231-1241.

HODGES, R.S. & MERRIFIELD, R.B. (1975b). Monitoring of solid phase peptide synthesis by an automated spectrophotometric picrate method. *Anal. Biochem.* **65**, 241-272.

HOPP, T.P. & WOODS, K.R. (1981). Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl. Acad. Sci. USA* **78**, 3824-3828.

JANIN, J. (1979). Surface and inside volumes in globular proteins. *Nature (London)* **277**, 491-492.

KARPLUS, P.A. & SCHULTZ, G.E. (1985). Prediction of chain flexibility in proteins. *Naturwissenschaften* **72**, s. 212.

LAEMMLI, U.K. (1970). Cleavage of structural proteins during the assembly of the head of Bacteriophage T4. *Nature (London)* **227**, 680-685.

LIU, F.T., ZINNECKER, M., HAMAOKA, T. & KATZ (1979). New procedures for preparation and isolation of conjugates of proteins and a synthetic copolymer of D-amino acids and immunochemical characterization of such conjugates. *Biochem. (U.S.A.)* **18**, 690-697.

MULLER, G., SHAPIRA, M. & ARNON, R. (1982). Anti-fluenza response achieved by immunization with a synthetic conjugate. *Proc. Natl. Acad. Sci. USA* **79**, 569-573.

NASH, A.A., LEUNG, K.-N. & WILDY, P. (1985). The T-cell-mediated response of mice to herpes simplex virus. In The Herpesviruses, Vol. 4, pp. 87-102. Edited by B. Roizman. Plenum Press, New York.

NAHMIAS, A.J., DANNENBARGER, J., WICKLIFFE, C. & MUTHER, J. (1981). Clinical aspects of infection with Herpes simplex viruses 1 and 2. In The Human Herpesviruses, pp. 3-10. Edited by A.J. Nahmias, Dowdle & Schinazi. North Holland:Elsevier.

NEURATH, A.R., STRICK, N., & KENT, S.B.M. (1985). Immune response to Hepatitis-B virus determinants coded by the Pre-S gene. In Vaccines '85, pp. 185-189. Edited by R.A. Lerner, R.M. Chanock, & F. Brown. Cold Spring Harbor Laboratory, New York, N.Y.

PARKER, J.M.R. & HODGES, R.S. (1985).I. Photoaffinity probes provide a general method to prepare synthetic peptide conjugates. *J. Prot. Chem.* **3**, 465-478.

PARKER, J.M.R., GUO, D., & HODGES, R.S. (1986). A new hydrophilicity scale derived from HPLC peptide retention data: Correlation of predicted surface residues with antigenicity and X-ray derived accessible sites. *Biochem. (U.S.A.)* **25**, 5425-5432.

SPEAR, P.G. (1976). Membrane proteins specified by herpes simplex viruses. I. Identification of four glycoprotein precursors and their products in type-1 infected cells. *J. Virol.* **17**, 991-1008.

SPEAR, P.G. (1985). Glycoproteins specified by herpes simplex viruses. In The Herpesviruses, Vol. 3, pp. 315-356. Edited by B. Roizman. Plenum Press, New York.

VOLLER, A., BIDWELL, D., HULDT, G. & ENGRALL, E. (1974). A microplate method of enzyme-linked immunosorbent assay and its application to malaria. *Bull. W.H.O.* **51**, 209-221.

WATSON, R.J., WEIS, J.H., SALSTROM, J.S. & ENQUIST, L.W. (1982). Herpes simplex virus type-1 glycoprotein D gene: Nucleotide sequence and expression in *Escherichia coli*. *Science* **218**, 381-384.

WATSON, R.J. (1983). DNA sequence of the herpes simplex virus type 2 glycoprotein D gene. *Gene* **26**, 307-312.

WENTWORTH, B.B. & FRENCH, L. (1969) Plaque assay of herpesvirus hominis on human embryonic fibroblasts. *Pro. Soc. Exp. Biol. Med.* **131**, 588-592.

WOROBEC, E.A., TANEJA, A.K., HODGES, R.S., & PARANCHYCH, W. (1983). Localization of the major antigenic determinant of EDP 208 Pili at the N-terminus of the pilus protein. *J. Bacteriol.* **153**, 955-961.

Table 2.1

Surfaceplot predicted surface sites and antigenic sites predicted using Hopp and Woods (1981) parameters

Surface site	Surfaceplot-predicted residues	Surfaceplot residues included in synthetic peptides	Hoppe and Woods-predicted residues
1	4-4	Yes‡	—
2	12-21†	Yes	11-13, 15-21*, 23
3	28-33*	No	29, 32-34
4	44-51*	No	—
5	66-66	Yes	64-66
6	73-78*	No	—
7	83-93†	Yes	84-91*
8	104-104	No	—
9	114-114	No	—
10	116-120*	No	119-120
11	122-122	Yes‡	—
12	124-124	Yes‡	—
13	130-140†	Yes	131,133
14	142-142	No	—
15	144-148*	No	145-147
16	160-162	No	—
17	172-172	No	169,172
18	174-175	No	—
19	184-189*	Yes	183-187
20	198-199	No	—
21	206-213*	Yes‡	—
22	224-229*	No	227-229
23	244-275†	Yes	256-259, 268-275*, 277-278
24	281-283	No	283-283
25	289-292	Yes‡	—
26	299-313†	Yes‡	—
27	322-322	Yes‡	—
28	324-324	No	—
29	340-351†	No	341-351†
30	354-354	No	—
31	356-365†	No	354-362

† Denotes predicted surface sequences of 10 or more residues and are referred to as major surface sites.

* Denotes predicted surface sequences of five or more residues and are referred to as minor surface sites.

‡ Denotes peptides synthesized that contained surface residues not predicted by Hopp & Woods program (Hopp & Woods, 1981).

Table 2.2

Sequences of HSV-1 glycoprotein D peptides synthesized

Sequence No.	Surface Site[a]	Peptide Sequence[b]
2-21	1, 2†	BB-G-G-(Y-A-L-A-D-A-S-L-K-M-A-D-P-N-R-F-R-G-K-D)-G-G-amide
12-21	2†	BB-G-G-(A-D-P-N-R-F-R-G-K-D)-G-G-amide
52-61	—	BB-G-G-(S-L-P-I-T-V-Y-Y-A-V)-G-G-amide
63-72	5	BB-G-G-(E-R-A-C-R-S-V-L-L-N)-G-G-amide
83-92	7†	BB-G-(G-A-S-E-D-V-R-K-Q-P)-G-G-amide
121-130	11,12	BB-G-G-(N-K-S-L-G-A-C-P-I-R)-G-G-amide
130-139	13†	BB-G-G-(R-T-Q-P-R-W-N-Y-Y-D)-G-G-amide
150-159	—	BB-G-(G-F-L-M-H-A-P-A-F-E)-G-G-amide
181-190	19*	BB-G-G-(L-E-H-R-A-K-G-S-C-K)-G-G-amide
206-215	21*	BB-G-G-(Q-A-Y-Q-Q-G-V-T-V-D)-G-G-amide
267-276	23†	BB-G-G-(E-L-A-P-E-D-P-E-D-S)-G-G-amide
288-297	25	BB-G-G-(P-Q-I-P-P-N-W-H-I-P)-G-G-amide
303-312	26†	BB-G-G-(A-T-P-Y-H-P-P-A-T-P)-G-G-amide
314-323	27	BB-G-G-(N-M-G-L-I-A-G-A-V-G)-G-G-amide

a Surface sites are defined in Table 2.1.

† Predicted surface regions of 10 or more residues.

* Predicted surface regions of 5 or more residues.

— Peptides synthesized as negative controls. These regions were predicted not to be on the surface of the protein.

b Peptides were synthesized based upon the HSV-1 glycoprotein D (Watson *et al.*, 1982).

Table 2.3

Sequences of peptides synthesized to investigate the effect of linker position and type of linker on the immune response to synthetic peptide conjugates

Sequence	Peptide sequence	Titer of antipeptide sera against[a]		
		Peptide-conjugate	Glycoprotein D	Virions
BB-G 12-21	BB-G-G-(A-D-P-N-R-F-R-G-K-D)-G-G-amide	4.0 (4.0)	4.0 (4.0)	3.1 (3.9)
BB-K 12-21	NAc-K(BB)-G-(A-D-P-N-R-F-R-G-K-D)-G-G-amide	4.0 (5.0)	2.0 (2.0)	1.8 (1.8)
12-21 BB-K	NAc-(A-D-P-N-R-F-R-G-K-D)-G-K(BB)-amide	4.8	3.0	2.0
C 12-21	C-(A-D-P-N-R-F-R-G-K-D)-G-G-amide	5.0 (5.0)	4.0 (4.0)	3.0 (3.6)
Y 12-21	Y-(A-D-P-N-R-F-R-G-K-D)-G-G-amide	5.0 (5.0)	4.0 (4.0)	3.0 (3.3)

[a] Titers from ELISA assays are expressed as negative logarithm of the reciprocal of the highest dilution of serum which still gave a positive response. A titer greater than 2 is considered significant. The value in brackets is a titer for the second rabbit. Antisera were tested against peptides conjugated to BSA, isolated glycoprotein D and HSV-1 virions.

Table 2.4

Results of antibody binding to HSV-1 virions and neutralization assays

Antibody to sequence[a]	Spanning surface site[b]	Antibody binding to HSV-1 virions	Titer of antipeptide sera against[c]			Neutralization[d]		
			Peptide conjugate	Glyco-protein D	HSV-1 virions	Yes or No	- Complement	+ Complement
2-21	1,2	Yes*	5.0 (5.0)	3.5 (3.5)	3.0 (3.3)	Yes*	1:25	1:200
12-21	2	Yes	4.0 (4.0)	4.0 (4.0)	3.1 (3.9)	Yes	1:25	1:25
121-130	11,12	Yes*	3.0 (3.2)	3.0 (3.0)	2.0 (2.9)	No		
206-215	21	Yes*	5.0 (5.0)	5.0 (5.0)	2.0 (2.6)	No		
267-276	23	Yes	5.0 (5.0)	3.5 (4.0)	3.2 (3.1)	Yes	1:50	1:50
288-297	25	Yes*	3.0 (3.3)	3.0 (3.0)	2.6 (3.2)	Yes*	1:25	1:25
303-312	26	Yes*	2.6 (5.0)	3.0 (3.5)	2.6 (2.6)	No		
314-323	27	Yes*	2.0 (3.0)	3.0 (3.5)	2.3 (2.7)	Yes*	1:25	1:25

[a] The amino acid sequences are shown in Table 2.2.

[b] Surface sites are defined in Table 2.1.

[c] The value in parentheses is the titer for the second rabbit (see Fig. 2.3). Titers in ELISA are expressed as the negative logarithm of the reciprocal of the highest dilution of serum which still gave a positive response. A titer greater than 2 is considered significant.

[d] Neutralization capacity was enhanced by the presence of 5% guinea-pig complement.

* Denotes synthetic peptides containing surface sites predicted by Surfaceplot and not predicted by the program of Hopp and Woods (1981) which generated antipeptide antisera which bound to HSV-1 virions or produced neutralizing antibodies.

Table 2.5

Sequences of synthetic peptides based on type-specific regions of glycoprotein D of herpes simplex virus

Sequence no. of glycoprotein D	Surface site[a]	Peptide sequence[b]	Titer of antipeptide sera against[c]		
			Peptide-conjugates	Glycoprotein D	HSV-1 virions
HSV1 38-47	4	BB-G-G-(Y-H-I-Q-A-G-L-P-D-P)G-G-amide	5.0 (5.0)	3.2 (3.2)	1.7 (1.7)
		* * *			
HSV2 38-47		Y-H-I-Q-P-S-L-E-D-P			
HSV1 83-92	7	BB-G-(G-A-S-E-D-V-R-K-Q-P)G-G-amide	5.0 (5.0)	3.6 (4.0)	1.7 (1.7)
		* * * * *			
HSV2 83-92		G-A-S-D-E-A-R-K-H-T			
HSV1 199-208	20,21	BB-G-G-(P-S-A-C-L-S-P-Q-A-Y)G-G-amide	3.0 (5.0)	2.6 (3.3)	1.7 (1.7)
		* * * *			
HSV2 199-208		P-A-A-C-L-T-S-K-A-Y			

[a] Surface sites are defined in Table 2.1.

[b] Asterisks denote non-conservative amino-acid changes between corresponding peptides for the glycoproteins D of HSV-1 and HSV-2. HSV-1 peptides were synthesized based upon the HSV-1 glycoprotein D sequence of Watson *et al.* 1982. HSV-2 peptides were from the HSV-2 glycoprotein D sequence of Watson *et al.* 1983

[c] The value in parenthesis is the titer for the second rabbit. A titer greater than 2 is considered significant, and a value of 1.7 represents a titer of <1 in 50.

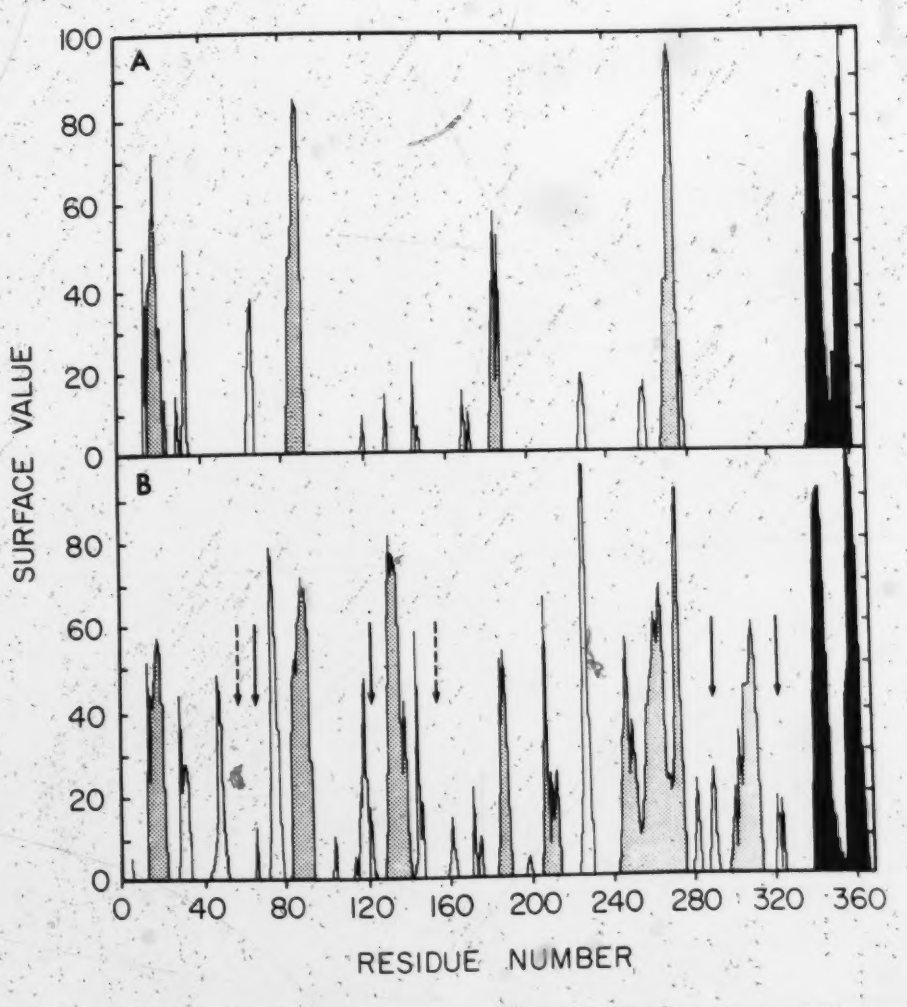


Fig. 2.1. Graphic representation of hydrophilicity (Panel A; Hopp and Woods, 1981) and Surfaceplot values (Panel B: Parker and Hodges, 1986) for glycoprotein D of HSV-1. Residues are numbered from the N- to the C-terminus. The surface value ordinate has been arbitrarily re-scaled such that the zero value corresponds to a cut-off of 25% above the mean of hydrophilicity/hydrophobicity values (Panel A) or the mean of exterior/interior values (Panel B). A surface value of 100 corresponds to the height of the highest peak in either case. The scale is therefore a relative one. Shaded peaks represent predicted surface regions of 5 or more contiguous residues (minor surface sites) or 10 or more contiguous residues (major surface sites) that were incorporated into synthetic peptides. The solid arrows denote regions where less than 5 contiguous residues were indicated on surface plot (Table 2.1). The dashed arrows represent regions predicted not to be on the surface. The solid peaks represent the two large surface areas corresponding to the hydrophilic C terminal cytoplasmic tail.

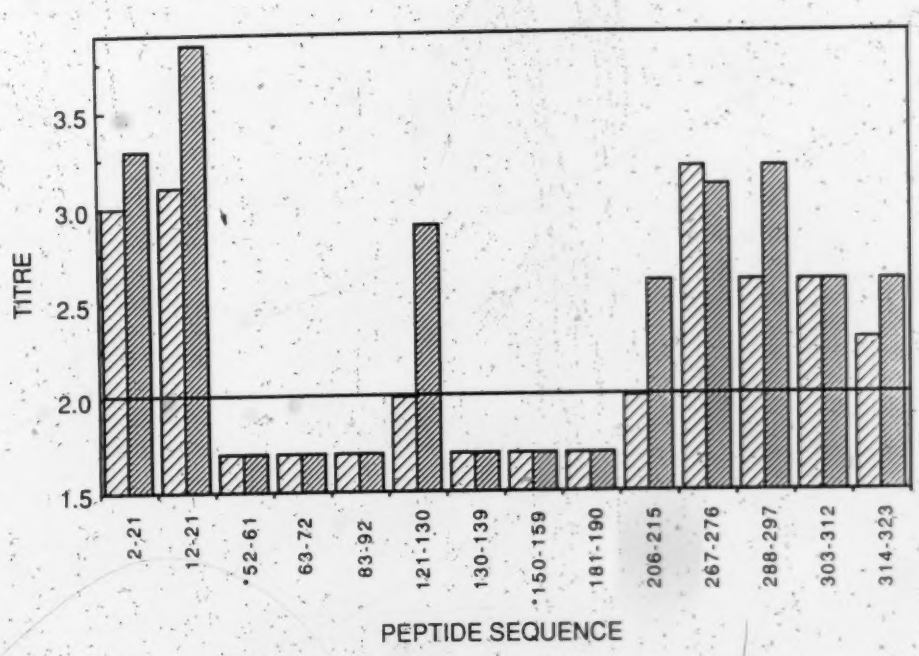


Fig. 2.2. Reactivity in ELISA of rabbit peptide antisera with HSV-virions (*i.e.* intact glycoprotein D). The asterisks indicate sequences which were selected and synthesized as negative controls. Serum with significant activity was judged to be that which reacted at dilutions of more than 1:100 (titer > 2.0). Antisera from two different rabbits were prepared for each peptide conjugates (differently shaded bars).

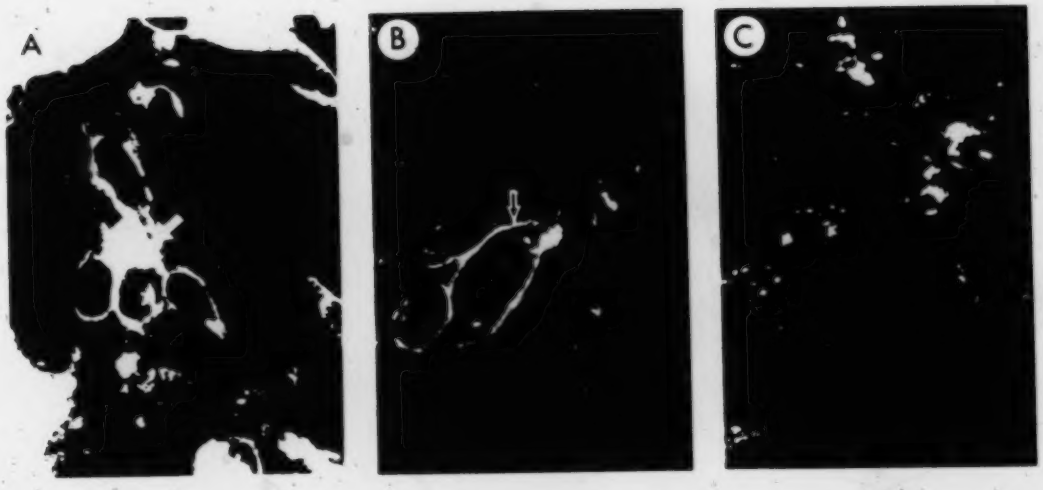


Fig. 2.3. Immunofluorescence of type-specific peptide antisera with HSV-1 and HSV-2 infected cells. Panel A shows the fluorescence pattern observed when HSV-1 infected cells are incubated with all three type-specific antisera. Note the pool of cytoplasmic fluorescence (arrow). Panel B shows the weak plasma membrane associated fluorescence observed when antisera to peptide sequence 38-47 is incubated with HSV-2 infected cells. Panel C shows the fluorescence observed when HSV-2 infected cells are incubated with antisera to peptide sequence 199-208. The photographic exposure for panels B and C was 5x longer than for panel A.

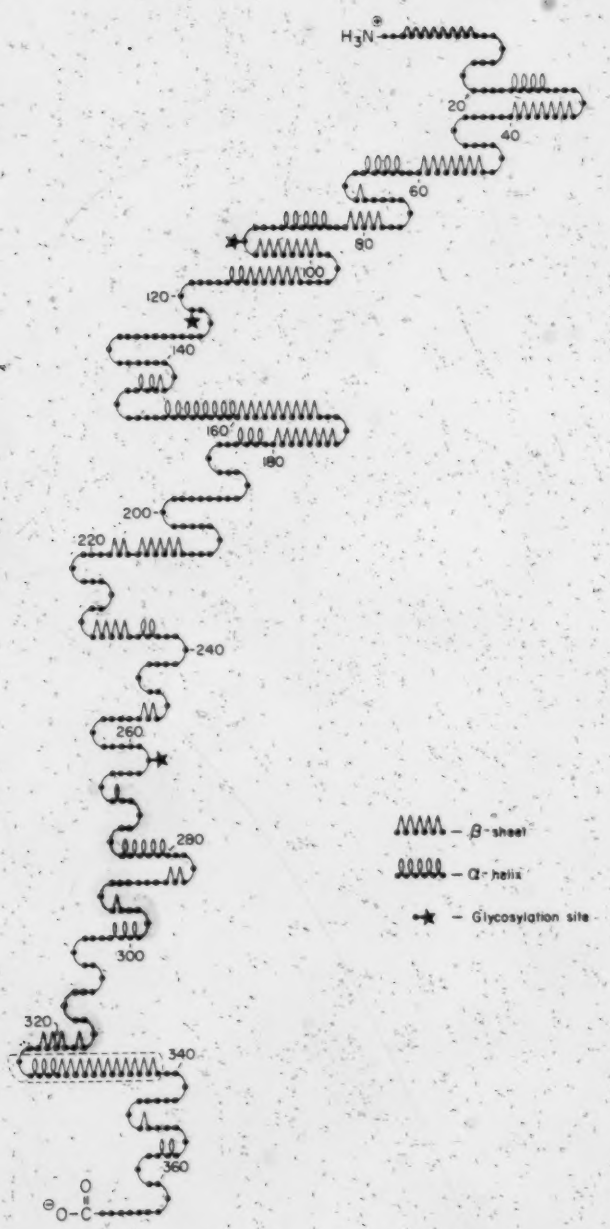


Fig. 2.4. Predicted secondary structure of glycoprotein D generated using a Chou-Fasman (1978) - based computer program. All major and minor linear surface plot predicted sites listed in Table 2.1 with the exception of 83-93 were found to correlate well with β-turns defined by Chou-Fasman analysis. All antigenic sites on the HSV-1 virion recognized by anti-peptide antibodies are outlined in dotted lines. The neutralization antigenic sites are also shaded. The proposed transmembrane region (residues 323-339) is outlined by a solid line.

CHAPTER 3

CRYSTAL STRUCTURES OF THE HELIX-LOOP-HELIX CALCIUM-BINDING PROTEINS[1]

LIST OF ABBREVIATIONS

TnC	troponin C
CaM	calmodulin
Parv	paryalbumin
ICaBP	intestinal calcium binding protein
MLCK	myosin light chain kinase
HLH	helix-loop-helix
EF-hand	a convenient mnemonic for the helix-loop-helix motif
rms	root mean square
NMR	nuclear magnetic resonance

[1] A version of this chapter has been published. Strynadka & James 1989. *Annu. Rev. Biochem.* **58**:951-98. Relevant material published subsequent to this paper is given in Chapter 7. This chapter is reproduced, with permission from the Annual Review of Biochemistry, Volume 58, C1989, by Annual Reviews Inc.

INTRODUCTION

Calcium ions have a number of diverse functions in biological systems, from biomineralization in bones, teeth and shells to a complex role as an intracellular messenger (1). The calcium binding proteins that have been subjected to high resolution crystal structure analyses fall into two general categories (Table 3.1). One group includes many extracellular enzymes and proteins that have enhanced thermal stability or resistance to proteolytic degradation as a result of binding Ca^{2+} ions. For some of the enzymes, Ca^{2+} may play an additional role in catalysis. The other group comprises a family of intracellular proteins that reversibly bind Ca^{2+} ions and thereby modulate the action of other proteins or enzymes. This second group is distinguished from the first in that its members have a common Ca^{2+}-binding motif consisting of two helices that flank a "loop" of 12 contiguous residues from which the oxygen ligands for the calcium ion are derived. The structure of a single helix-loop-helix (HLH) motif has been likened to an index finger (E-helix), a curled second finger (the loop) and a thumb (F-helix) of a right hand, so that the term "EF-hand" has been widely applied to describe such a Ca^{2+}-binding site (29). A second distinguishing characteristic of the HLH family is the fact that the Ca^{2+}-binding motifs occur in intimately linked pairs (Table 3.1). With the exception of thermolysin, all other proteins in Table 3.1 contain only one calcium binding site that has no associated helices. Thermolysin has a double site (two Ca^{2+} ions separated by 3.8 Å) in which one Ca^{2+} ion is 6 coordinate and the other 7 coordinate (13,14).

There have been a number of reviews concerning various structural aspects of calcium-binding to small molecules and to proteins (44-47). Subsequently, additional high resolution crystal structure analyses or refinements of earlier structures have been completed. In writing this review, we have been extremely fortunate to have had access to the refined atomic coordinates of bovine brain calmodulin, CaM (38), turkey skeletal muscle troponin C, TnC (41), chicken skeletal muscle TnC (43), carp parvalbumin, Parv (31) and bovine intestinal calcium binding protein, ICaBP (34). This has allowed us to make direct structural comparisons among the members of this family.

There are several new deductions and generalizations regarding Ca^{2+}-binding sites that can be made from an analysis of the data base of new structures (Table 3.1). The majority of Ca^{2+}-binding sites contribute seven oxygen ligands to the metal ion. There are only three instances in which the Ca^{2+} coordination number is six. For those with seven ligands, the oxygen atoms are located approximately at the apices of a pentagonal bipyramid ~ 2.4 Å from the central Ca^{2+} ion. Not all of the seven ligands are from the protein; many of the binding sites have one or more water molecules in the coordination sphere of the Ca^{2+} ion. The HLH family has a very characteristic and consistent deviation of one of the 5 ligands from the pentagonal plane. This deviation may be important in providing for both Ca^{2+} and Mg^{2+} binding.

The calcium-binding sites of the HLH proteins involve a segment of polypeptide chain with 12 contiguous residues, whereas the sites in the other proteins are 'discontinuous' with the Ca^{2+} ligands coming from residues on widely separated segments of polypeptide chain (Table 3.1). There are two exceptions in this last group; bovine trypsin (2) and α-lactalbumin (24) have continuous calcium-binding segments of 11 and 10 residues, respectively.

Each of the 12 residues of the loop in an HLH motif plays an important role in defining the structure of the calcium binding site. Invariably, five of them are involved directly in providing oxygen ligands to the Ca^{2+} ion. The other residues provide hydrogen bonding via main chain NH groups to stabilize the geometry of the loop required for Ca^{2+} binding.

The majority of the helical segments in the HLH proteins display normal α-helical geometry with the expected n to n+4 hydrogen bonding pattern. However, there are several helices (one in each protein) that are considerably bent or distorted, allowing more favorable hydrophobic interactions. Helix crossing angles have often been used as indicators of the overall molecular conformations (34,36,39,44). However, detailed comparisons of the several structures show that these parameters are not particularly

informative indicators of conformational differences. The interhelical packing between HLH structural units is strongly conserved. Comparison of the Ca^{2+}-free N-terminal domain of TnC with its C-terminal domain indicates that conformational changes that occur on Ca^{2+}- binding probably involve coupled movements of a pair of helices (48). Such movements on Ca^{2+} binding would expose an extensive hydrophobic patch on the molecular surface. A variety of biophysical studies on TnC and CaM support this model for the Ca^{2+}-mediated conformational change.

FUNCTIONAL OVERVIEW

The helix-loop-helix Ca^{2+}-binding proteins are a family of highly analogous, intracellular proteins whose activities are regulated by the Ca^{2+}-binding event. Binding Ca^{2+} to these proteins induces in them a conformational change that is subsequently transmitted to their respective target molecules. For this reason they are often termed "Ca^{2+}-modulated". The two most well characterized examples of the family in terms of structure and function are TnC and CaM.

The cellular functions of two other members of the HLH family are less clear. Although the strong primary and tertiary structural homology of Parv and ICaBP warrant their inclusion in the TnC/CaM family, the term "Ca^{2+}-modulated" cannot be applied to them. There are as yet no firm biochemical data that would indicate whether Parv and ICaBP directly modulate the activity of any specific secondary target molecules. Most evidence thusfar suggests a more passive role for Parv and ICaBP, in that binding of calcium to these molecules aids in the regulation of calcium concentrations in the cell.

Because protein structure and function are such an inseparable marriage, one cannot fully discuss one aspect without referring to the other. For this reason, a brief summary of the current views of the functionality of these four HLH proteins is presented.

Troponin C

TnC is a key player in the Ca^{2+}-mediated regulation of muscle contraction. The

cross-bridge model proposes that the S1 heads of myosin in the thick filaments cyclically attach and detach from the actin-containing thin filaments (49). With myosin bound to actin, the complex forms a Mg^{2+}-activated ATPase, actomyosin. As ATP is hydrolyzed, the two filaments slide past one another and the contractile force is generated. Regulation of the contraction/relaxation cycle in skeletal muscle is Ca^{2+}-mediated at the level of the thin filament through the protein complex of troponin and tropomyosin (50-52). TnC is the acidic (pI 4.25) 18,000 dalton Ca^{2+}-binding component of troponin, a tripartite complex consisting of TnC, TnI and TnT (53,54). Troponin binds at regular intervals along the actin polymer of the thin filament. TnT is primarily responsible for binding the complex to the long coiled-coil molecule of tropomyosin. TnI inhibits the Mg^{2+}-activated ATPase of actomyosin (53). TnC binds only to specific regions of TnI, and the strength of this interaction is increased in the presence of Ca^{2+} (55-60). Several studies suggest that TnI binding occurs on hydrophobic surfaces of TnC that become accessible when Ca^{2+} binds (61-64). Whatever the case, Ca^{2+} binding to TnC induces a conformational change that directly affects its interaction with TnI (65,66). This results in a change in the disposition of the troponin/tropomyosin complex relative to actin such that either a steric hindrance to the approach of myosin heads to actin is removed (49,67,68) or the release of inorganic phosphate (Pi) from the actin/myosin head/ADP:Pi complex, normally blocked in the relaxed state, is allowed (69).

Skeletal TnC contains two high affinity Ca^{2+}-binding sites ($K_d \sim 10^{-7}$ M). These sites also bind Mg^{2+} competitively ($K_d \sim 10^{-3}$ M) and are therefore called the Ca^{2+}-Mg^{2+} sites (54,63,70). Skeletal TnC also contains two sites of lower Ca^{2+} affinity ($K_d \sim 10^{-5}$ M) that are essentially specific for Ca^{2+} at physiological Mg^{2+} concentrations (71).

Calmodulin

CaM is an intracellular calcium receptor found in all eukaryotic cells from yeast to higher mammals (72). It is a small, acidic protein of molecular weight 16,700 whose cellular function encompasses the Ca^{2+}-mediated activation of a number of different

intracellular enzymes (73) including phosphodiesterase (74,75), myosin light chain kinase (MLCK) (76,77), calcineurin (78), erythrocyte Ca^{2+}-ATPase (79,80), brain adenylate cyclase (81,82), phosphorylase kinase (83) and nicotinamide dinucleotide kinase (84). The Ca^{2+}-saturated molecule is the active form of CaM. As in the case of TnC, conformational changes produced by Ca^{2+} binding are thought to result in exposure of hydrophobic surfaces with which target enzymes or inhibitory drugs interact (85-95).

CaM has four binding sites with association constants for Ca^{2+} falling within one order of magnitude of one another. It is generally accepted that the four Ca^{2+}-specific sites of CaM can be divided into the lower affinity sites I and II with $K_d \sim 10^{-5}$ M and sites III and IV with slightly higher affinity ($K_d \sim 10^{-6}$M) (89,96-100). Biochemical characterization of CaM is extensive, and has been discussed recently elsewhere (85,101,102).

Parvalbumin

Parvalbumins are a large group of Ca^{2+}-binding proteins that have been isolated mainly from fast twitch muscles of fishes, amphibians and mammals (103). Members of the family exhibit a variety of molecular weights and isoelectric points (pI). Carp Parv is small (11,000 daltons) and acidic (pI ~ 4.25). Although their function is not defined, several kinetic and equilibrium Ca^{2+}/Mg^{2+}-binding studies suggest that they may be involved in the relaxation event following muscle contraction (104-107). It has been proposed that Parv is usually in the Mg^{2+} bound form in muscle. At physiological levels of Mg^{2+} (1 mM) and K^+ (80 mM), and at levels of Ca^{2+} corresponding to those of resting muscle (~ 10^{-8}M), Parv binds two Mg^{2+} ions and no Ca^{2+}. When Ca^{2+} is released into the cell following a nerve impulse, the Ca^{2+} binds first to TnC and CaM, presumably because of the very slow off rate of Mg^{2+} from Parv. Upon muscle relaxation, Parv could take up the Ca^{2+} released by TnC and CaM, as Mg^{2+} is released. Therefore it could act as a calcium buffer, by quickly reducing the calcium concentration so contraction is not reinitiated. That parvalbumins are found in greatest quantities in fast twitch muscles may support this conclusion (108). It should be noted that parvalbumins are essentially skeletal

muscle proteins; they are usually not found in cardiac or smooth muscle and so are not indispensable components of the contractile mechanism.

Parvalbumins bind Ca^{2+} very tightly with impressive K_d's of up to 10^{-9} M (106,108-110). Mg^{2+}, Na^+ and K^+ ions compete with Ca^{2+} for the metal ion binding sites of Parv only at high concentrations (109,111). Under normal physiological conditions, Parv is not found in the apo metal-free state because of the exceptionally high affinity for this ion.

Intestinal Calcium Binding Protein

The vitamin-D dependent ICaBP is also known as calbindin 9K. It is a soluble protein located primarily in the cytoplasm of the absorptive cells of mammalian intestinal tissue (112). As with Parv, the function of ICaBP is not certain. Various studies support its role as a Ca^{2+} buffer for vitamin D stimulated Ca^{2+} absorption (113,114); others suggest it may be an aqueous intracellular Ca^{2+} transporter (115). Preliminary biophysical studies indicate that ICaBP does not undergo a large conformational change upon Ca^{2+} binding (116,117).

ICaBP binds two Ca^{2+} ions with moderate affinity ($K_d \sim 10^{-6}$ M to 10^{-8} M) (118-121). It does not appear to bind Mg^{2+} at physiological concentrations.

STRUCTURAL OVERVIEW

Primary Structures

Table 3.2 presents the aligned amino acid sequences of chicken skeletal TnC, bovine brain CaM, carp Parv and bovine vitamin-D dependent ICaBP. By far the greatest sequence identity occurs between CaM and TnC (51% identity). Parv and ICaBP show considerably less analogy with each other and with other members of the family.

In Table 3.2 the helices of each HLH unit are labelled alphabetically and the loops are denoted with Roman numerals. The nomenclature of the Ca^{2+}-binding loop derives from the original octahedral description of the Ca^{2+}-coordinating ligand geometry (29). In

that description, six residues in positions 1(X), 3(Y), 5(Z), 7(-Y), 9(-X) and 12(-Z) provided oxygen ligands to the Ca^{2+} ion. In all known HLH units, the residue at -Z is a glutamate; it contributes both of its side chain oxygen atoms to the metal ion coordination. This bidentate coordination at -Z along with the oxygen ligands at Y, Z and -Y define an approximately planar pentagonal arrangement. The invariant aspartate at X contributes one oxygen to the pyramidal apex, the opposite apex (-X) is occupied generally by a water molecule (39). Only in loop I of Parv is there a residue in position 9 that has a side chain sufficiently long to coordinate directly to the Ca^{2+}. Therefore, since there are seven oxygen ligands, the true Ca^{2+}-coordination is pentagonal bipyramidal, not octahedral as previously described. In spite of this we will use the established octahedral nomenclature given in Table 3.2. Also shown in Table 3.2 is the three-residue overlap between the Ca^{2+}-binding loops (positions 10, 11 and 12) and the following helices.

All four proteins have N-terminal extensions of variable lengths that precede the first HLH folding unit (TnC, 15 residues; CaM, 5 residues; Parv, 39 residues and ICaBP, 2 residues). The linking peptides between HLH domains in TnC and CaM are five amino acids long. Those of Parv and ICaBP are 8 and 10 residues long, respectively.

The HLH proteins are characterized by relatively high percentages of acidic residues (TnC, 29%; CaM, 25%; Parv, 18% and ICaBP, 23%). There are also a number of conserved hydrophobic amino acids, many of which are aromatic phenylalanine residues. None of the four proteins has a tryptophan; chicken TnC has no tyrosine; bovine brain CaM has no cysteine; carp Parv has no methionine, tyrosine or proline; bovine ICaBP has no methionine, arginine, cysteine or histidine.

The conserved sequences of the Ca^{2+}-binding loops and their flanking helices have allowed for the derivation of a consensus sequence (34,64). Successful prediction of HLH binding sites in newly determined protein sequences has been possible with these sequences (128-130).

The chicken sequence of TnC originally determined by Wilkinson (122) has been

completed by DNA sequencing methods (123). DNA sequencing methods have corrected the original report that in CaM residue 129 was an asparagine (124). It has subsequently been confirmed as aspartic acid (125,131). The crystal structure of CaM that has been reported (38) still retains the originally assigned asparagine at position 129.

Tertiary Structure

CRYSTALLOGRAPHIC ANALYSES

In this section we review briefly the reliability of the five crystal structure determinations. Table 3.3 presents a summary of the crystallographic data. All five structures have been refined to final R factors of 0.155 to 0.187. These values are indicative of well refined crystal structures at medium resolution.

Both avian skeletal TnC structures were refined at 2.0 Å resolution and are comparable determinations (41,43). Analysis of the coordinates by a least-square minimization procedure reveals that the two molecules are very similar. For the 628 common main-chain atoms, the root mean square (rms) difference is 0.29 Å; for the 1210 common atoms of whole TnC, the rms difference is 0.66 Å. The larger number is primarily a reflection of the greater inaccuracy in coordinates for the more mobile side chains, like glutamates and lysines. Other than these residues, there are only two major conformational differences in the side chain positions between turkey and chicken TnC at Phe105 and Ile115. Overall, the differences in these two completely independent determinations of similar molecules are minimal and compare favorably to results on dual determinations of other protein structures [e.g. bovine trypsin (135) and bovine chymotrypsin (136-138)]. Four residues were not seen in the final electron density map of turkey TnC, the first two, the last one and the side chain of Glu67.

The CaM structure has been refined at 2.2 Å resolution (38). The positions of five of its residues were not defined also due, most likely, to disorder. The refinement of the Ca^{2+} and Cd^{2+} complexes of carp Parv have been done at the highest resolution, 1.6 Å (31). Since these two structures were isomorphous and have essentially identical tertiary

structures, we have included only the results for the Ca^{2+} Parv in the present discussion. The previously reported structures of Parv were of lower resolution and not completely refined (29,30). Finally, ICaBP has been refined at 2.1 Å (34).

Solvent molecules are an integral component of protein structure. Analysis of the results from high resolution (better than 2.0 Å) protein structures indicates that the ratio of the number of ordered solvent sites to the number of residues in the protein is roughly 1:1. This would indicate (see Table 3.3) that for chicken TnC, CaM, Parv and ICaBP, the description of ordered solvent may still be incomplete, thus precluding detailed comparisons. Nonetheless, it is likely that all the strongly bound solvent molecules have been selected, and among them there are many conserved solvent sites common to the five protein structures.

Table 3.3 also gives the rms deviations of each of the model structures from ideal stereochemistry. The size of these numbers, in conjunction with the amount of data included in the refinement and the value of the agreement factor R, give a general indication of how complete the structure refinements have been. Chicken skeletal muscle TnC (43) has rms deviations that are approximately twice those of the other four structures. Other parameters are comparable among all of them.

Estimates of the error in the refined atomic coordinates range from 0.15 Å to 0.30 Å. The lower estimates for CaM and chicken TnC are overly optimistic. The low value of the coordinate errors (0.15 Å) for the Parv structure reflects the much higher resolution of that study. For turkey TnC, the more conservative estimate for the coordinate accuracy was determined from a σ_A plot (139).

OVERALL ARCHITECTURE

Analysis of the crystallographic structures of TnC, CaM, Parv and ICaBP reveals a conserved structural arrangement of each Ca^{2+}-binding domain. It consists of a pair of HLH structural motifs joined by linker peptides of 5 to 10 residues in length. Figure 3.1 shows a diagrammatic representation of one domain (the HLH units are generically denoted

as helix A-loopI-helix B and helix C-loop II-helix D). The helices are approximately 10 to 12 residues long and they flank a 12 residue loop. The HLH structures are related to each other by an approximate intradomain 2-fold rotation axis centrally located between the two loops (Figure 3.1). Single HLH units that bind calcium have not been observed thusfar. All Ca^{2+}-binding domains analyzed to date involve a pair of HLH motifs, even though in some cases one of the sites may have lost its Ca^{2+}-binding ability [cardiac TnC (140); crayfish TnC (141)].

In all determined structures, the two HLH motifs are intimately associated. The inter-motif interactions occur via the loops and the helices. The two loops of adjacent HLH motifs interact via two antiparallel β-sheet hydrogen bonds. This is the only β-sheet secondary structure in the molecules. The nature and arrangement of the helices also play a critical role in the interaction of HLH motifs. All four helices of the domain are amphipathic. They pack with their hydrophobic faces inward, away from solvent, forming a central core that consists of a number of intra- and inter-helical hydrophobic interactions. On their solvent-exposed faces, charged residues form favorable electrostatic interactions by binding metal ions, bulk solvent, or more rarely, by forming intramolecular ion pairs. The Ca^{2+}-binding domain can be thought of as cup-shaped (Figure 3.1); the inside of the cup is lined with relatively exposed hydrophobic residues, the bottom of the cup contains the Ca^{2+} binding loops, and the outside and rim of the cup is comprised of charged, hydrophilic residues. The similarity in tertiary structure of the Ca^{2+}-filled domains can be appreciated by comparing them pairwise using a least-squares procedure (Table 3.4). The comparison excluded the ICaBP structure because it has major conformational differences with the other domains. The smallest rms difference is between the C-terminal domain of TnC and the C-terminal domain of CaM. On the other hand, the largest rms difference is between TnC-C and CaM-N. While the Ca^{2+}-binding domain is conserved in both sequence and architecture, significant structural differences among the four molecules arise mainly from modifications to the linker regions that bind adjacent HLH motifs together,

and from structural additions to the N-termini of the Ca^{2+} binding domains.

Troponin C — TnC is a 70 Å long dumbbell-shaped molecule. The globular N- and C-terminal domains are connected by a 31 residue α-helix (Figure 3.2a). The domains have mean radii of approximately 17 Å; their centres are separated by 44 Å (41). Each domain of TnC consists of a pair of HLH motifs. The pair in the N-terminal domain are helix A-loop I-helix B and helix C-loop II-helix D. The N-terminal domain has an additional helical structure, the N-helix, that is unique to the TnCs. Helix E-loop III-helix F and helix G-loop IV-helix H form the EF-hand pair in the C-terminal domain. The central connecting peptide between the two domains is formed by the continuous helical sweep from the D-helix, through the D-E helical linker region, to the E-Helix (Figure 2a and Table 3.2).

In both X-ray crystallographic structures of chicken and turkey TnC discussed here (41,43), the C-terminal high affinity sites (III and IV) are occupied by metal ions, whereas the two Ca^{2+}-specific low affinity sites (I and II) in the N-terminal domain have no bound Ca^{2+} ions. This lack of Ca^{2+}-binding is attributed to the low pH at which the crystals were obtained (Table 3.3). Comparison of the structures of the pair of HLH motifs in the N- and C-terminal domains indicates marked tertiary structural difference between the Ca^{2+}-free and Ca^{2+}-bound forms.

Calmodulin — From the very high sequence identity and similar biochemical properties of TnC and CaM one would expect extensive structural similarities between the two molecules (142). Comparison of Figures 3.2a and 3.2b shows that they do exhibit similar portraits. There are three major structural differences between TnC and CaM. Both the N- and C-terminal domains of CaM are in the Ca^{2+}-bound form; the N-terminal domain of TnC is metal free. The rms difference for the main-chain atom comparison of the N- and C-terminal domains of CaM (0.751 Å, Table 3.4) is indicative of the very similar conformations. The N-terminal helical arm of TnC has no counterpart in CaM. There is a

three residue deletion in the central D/E helical linker of CaM (see also Table 3.2). As a result of the deletion, CaM is approximately 5 Å shorter than TnC, with a length of 65 Å (38) and the N- and C-terminal domains are oriented differently. In Figure 3.2a and 3.2b the C-terminal domains of TnC and CaM are displayed from the same vantage point. It can be seen that relative to the fixed C-termini, the orientation of the N-terminal domain of TnC differs by ~ 60° to that of CaM.

The different lengths and resulting orientation of domains in TnC and CaM may play a role in differentiating their function. A recent study has shown that a three residue insertion into the central helix of CaM causes a marked decrease in the activation of some of the molecule's target enzymes (143).

The dumbbell shapes of TnC and CaM were totally unexpected (144). The refined atomic coordinates from the crystal structure studies show interatom pair distribution functions [P(r)] that are bimodal. The first peak corresponds to the mean radius of the Ca^{2+}-binding domains and the second peak, to the mean distance between domains (41,145,146). In addition, several small angle X-ray scattering studies (SAXS) have evaluated the shapes of these molecules in solution (145-148). These studies showed P(r) distributions that are similar to that of the X-ray crystal structures, but with the minimum between the peaks not so pronounced. The maximum lengths of both CaM and TnC determined in solution agree well with the lengths from the single crystal X-ray studies. The results are thus consistent with the presence of a population of molecules in solution that have the extended dumbbell shape. In order to explain the larger number of intermediate length pair distances (between 25 and 40 Å) on the solution P(r) function, two alternatives have been put forward. One of them proposes a population of molecules in solution with a bent central helix (145). A bend of ~ 65° produces good agreement with the values of the P(r) function in solution. The alternative explanation involves bound water molecules between the domains. This hydration layer would tend to raise the minimum in the interatomic distance plot (at ~ 30 Å), thereby accounting for the shape of

the experimental curves (146). Most likely both explanations are valid. No clearcut experiments that would differentiate between these alternatives have yet been done.

Parvalbumin — Unlike the dumbbell-shaped TnC and CaM, Parv is a globular molecule (Figure 3.2c) that can be described as a prolate ellipsoid with dimensions of 36 x 30 x 30 Å (30). Parv contains six helices A through F (Table 3.2). Helices C and D flank loop I, E and F flank loop II. Collectively, they constitute the familiar cup-shaped Ca^{2+}-binding domain. In addition, Parv has a 39 residue N-terminal extension comprised of two helices, A and B, that flank an eight residue loop. It has been proposed that this region is a defunct Ca^{2+}-binding site that has arisen by gene triplication (149). Unlike TnC and CaM, whose N- and C-terminal Ca^{2+}-binding domains are linked by a long central helix, the single N-terminal HLH motif of Parv folds over, packing its hydrophobic face into the hydrophobic cup formed by helices C,D,E, and F of the Ca^{2+}-binding domain. As a result, Parv has none of the exposed hydrophobic surfaces which are thought to be the sites of target molecule binding in TnC and CaM. Thus, Parv depicts a globular shape, with a buried central core of hydrophobic residues and an outer shell of hydrophilic ones.

Intestinal Calcium Binding Protein — ICaBP has 75 amino acid residues and is the smallest of the known HLH proteins. It consists of only four helices (A, B, C and D) and two loops (I and II) that make up the two HLH motifs of the Ca^{2+}-binding domain (Table 3.2 and Figure 3.2d). ICaBP is also an approximate prolate ellipsoid with dimensions of 30 Å long and 25 Å in diameter (34). The most distinguishing features of ICaBP are the altered Ca^{2+}-binding loop I [a result of a two residue insertion (Table 3.2)], and the ten residue linking peptide that joins the two HLH structural units between helices B and C. This extended peptide linker has hydrophobic residues that contribute their side chains to the hydrophobic interior of the cup formed by the inner surfaces of helices A to D. ICaBP, like Parv, has no significant hydrophobic patches on the surface that might serve as interaction sites for target molecules.

THE CA^{2+} COORDINATION

The coordination of Ca^{2+} ions in the calcium binding proteins that have been studied crystallographically has often been referred to as octahedral, thereby implying six ligands. The more highly refined crystal structures now show that there are seven oxygen atoms at an average distance of 2.4 Å from the Ca^{2+} ion. Five of them (in an approximate pentagonal arrangement) lie close to a common plane that includes the central Ca^{2+} ion. The vector joining the other two oxygen atoms passes near to the Ca^{2+} ion and is approximately perpendicular to this pentagonal plane (Figure 3.3). This defines the coordination geometry as pentagonal bipyramidal. Figure 3.4 clearly shows that this is the case for all of the HLH proteins. In fact, almost all of the presently determined and highly-refined crystal structures of proteins that bind Ca^{2+} ions do so in this seven coordinate fashion (Table 3.1). The three exceptions seem to be bovine trypsin (2) in which the Ca^{2+} is six coordinate, thermolysin which has a six coordinate and a seven coordinate double site (14) and staphylococcal nuclease, details of which have not been fully published (20,47).

Many Ca^{2+}-binding motifs have been observed in proteins (Table 3.1). Some have all seven ligands originating from atoms in the protein (parv, subtilisin). Others, including most members of the EF hand family, have one water molecule in their ligand sphere; some have two water ligands (e.g., SGT) and others have three water ligands (e.g., site 3 of thermolysin). One extraordinary calcium binding geometry is that from rhizopus chinensis aspartic proteinase (150); it has a single protein ligand (a main chain carbonyl oxygen atom) and six water molecules that complete the seven coordinate pentagonal bipyramidal geometry.

Figure 3.3a shows the idealized pentagonal bipyramid with its seven vertices at a distance of 2.40 Å from the central Ca^{2+} ion. Superimposed upon that are the oxygen ligands of loop III from the refined CaM structure (38). The equatorial ligands are Asp95 $O^{\delta 2}$, Asn97 $O^{\delta 1}$, Tyr99 O and the two carboxyl group oxygen atoms of Glu104. The bidentate coordination of this side chain forces the oxygen atoms at positions Y and -Y

towards the oxygen atom of the ligand at Z. We have done a least-squares fitting of the seven oxygen ligands of each of the HLH calcium binding loops to the idealized pentagonal bipyramid. The rms deviations from the ideal for the ten refined loops range from 0.36 Å to 0.50 Å.

The HLH Ca^{2+}-binding proteins depart from strict coplanarity of the Ca^{2+} ion and the five oxygen ligands that determine the equatorial pentagonal plane (Figure 3.3b). The deviation is consistently exhibited by all of the metal binding sites in these proteins. The glutamate in position 12 (-Z) has one carboxylate oxygen that is in the plane but the carboxylate group has been rotated so that the other oxygen atom is out of the plane by ~ 1.2 Å. Thus, the pentagonal arrangement of oxygen ligands has an envelope pucker that is reminiscent of one of the commonly observed conformations of a ribose ring (151).

Figure 3.3b also indicates that the carboxylate or amide planes of the ligands in positions X, Y and Z are not directed towards the central Ca^{2+} ion. Rather, the Ca^{2+} is displaced by ~ 1.0 to 1.5 Å out of each of these planes (Figure 3.4b-f). This is not the case for most of the Ca^{2+}-binding proteins outside the HLH family (Table 3.1). For these, the metal ion lies close to the carboxylate or amide planes.

It is tempting to speculate that the coordination geometry of the HLH proteins is designed to accommodate, separately, both Ca^{2+} and Mg^{2+} ions. The average ligand distance for Mg^{2+} is shorter, ~ 2.1 Å (152) and its coordination number is predominantly six. One of the main determinants of the Ca^{2+} seven-fold coordination is the bidentate nature of the glutamate carboxylate and the conformation of this side chain. In order to adapt this site so that it would bind Mg^{2+}, subtle changes in the torsional angles χ_2 and χ_3 would rotate the carboxylate so that its plane would be approximately perpendicular to the equatorial plane. In this configuration only one of the oxygen atoms of the carboxylate would still be on the equatorial plane and coordinating to the metal ion, allowing the other five coordinating ligands to cluster more closely around a smaller central Mg^{2+} ion with the appropriate octahedral coordination. This could be achieved through rotations of ~ 30°

about χ_1 in the residues at positions X, Y and Z, in such a way that the carboxylate or amide planes of these residues would now be directed towards the Mg^{2+} ion. This subtlety may be required for those sites that adjust to both Mg^{2+} and Ca^{2+} ions.

The seven-coordinate geometry seen for the Ca^{2+} ions holds equally well for the Cd^{2+} ions in the refined Cd parvalbumin (31). The average $Cd^{2+}\cdots O$ distance in that structure was also 2.40 Å. The seven coordinating oxygen atoms that have been confirmed in the Cd-Parv refinement agree well with the ^{113}Cd NMR work (153), a study that indicated that the coordination number of the metal had to be greater than six for both metal binding sites of parvalbumin.

Comparison of Loop Conformations — In order to assess the similarity of conformation of the loops in these several proteins (Figure 3.4), we have superimposed the 48 main-chain atoms of each loop (4 x 12 residues) with one another and in pairs (96 atoms in each pair) via a least-squares procedure. The results of the comparisons of the pairs of loops are summarized in Table 3.5. For the C-terminal domain of TnC for both domains of CaM and for Parv, the rms deviations are approximately 0.50 Å. The pair of loops in ICaBP have a slightly different conformation. This is due to the altered binding site I. For this protein only those residues of loop I deemed to be topologically similar to the residues of the common loop were used in the comparisons (Gln22 to Glu27).

The comparisons of the individual loops among themselves showed two important features. Firstly, the rms deviations for the 36 comparisons ranged from 0.28 Å (TnC IV vs Parv II) to 0.68 Å (CaM III vs CaM IV). These individual values are similar in magnitude to the values that result when the loops were compared as pairs (Table 3.5). Therefore, in the presence of Ca^{2+}, the two β strands within the loops interact with one another in a common fashion (twist angle is ~ -34°). Secondly, comparisons of loops I with other loops I or III and loops II with other loops II or IV show significantly smaller rms deviations than the values resulting from comparisons of loop I with loops II or IV and loops III with loops II or IV. Thus, loop I is structurally more similar to loop III and loop

II is more similar to loop IV. Since all of the loop regions have highly homologous sequences (Table 3.2), the reason for the structural differentiation between I, III and II, IV must arise from the nature of the flanking helices. Indeed, the greatest sequence homology is between helices flanking loops I and III and those flanking loops II and IV. This EF-hand homology was first noted by Weeds and McLachlan (154). In particular, the conserved pair of aromatic residues on helices A and E, preceding loops I and III is not present on the helices C and G, that precede loops II and IV. Similarly, the aromatic residues on helices D and H that follow loops II and IV are not on helices B and F.

The Standard Ca^{2+}-Binding Loop — Perhaps the most convenient way to discuss the conformation of a generic Ca^{2+}-binding loop is to proceed from the N-terminus to the C-terminus and to discuss each residue in turn with regards its contribution to the metal coordination and to the loop conformation. The values of ϕ,ψ and χ_n that are given for each position are the averages for the 11 loops that have Ca^{2+} bound in the five structures we are discussing. The spread of values of conformational angles is small for most residues of the loops. The variant loop I of ICaBP and the Ca^{2+}-free loops of TnC will be discussed separately.

Position 1 ($\langle\phi,\psi\rangle$ = -75°, 84°; $\langle\chi_1,\chi_2\rangle$ = -176°, 21°).

The first residue of the loop is invariant; it is an aspartate. The incoming helix is quite regular up to this residue but it finishes here (ψ is +84°). The ϕ,ψ angles are characteristic of those of the middle residue of a γ turn (155) in which there is a main-chain hydrogen bond between the first and third residues (156). This hydrogen bond, from the NH of the residue in position 2, to the CO of the last residue in the helix (position -1) is present in all loops. The turn does not reverse the chain direction, but is the secondary structural feature that terminates the incoming helix and initiates the loop.

The carboxylate group of the side chain plays a prominent role in Ca^{2+} binding (position X in Table 3.2) and in providing a focus around which the first six residues of the loop fold. One of the oxygen atoms of the carboxylate is a direct Ca^{2+} ion ligand

(Figure 3.4) and it is one of the pyramidal apices of the pentagonal bipyramid. The other oxygen atom is the recipient of a strong hydrogen bond from the main-chain NH of the conserved glycine in position 6 (Figure 3.5). It also forms a hydrogen bond to a conserved water. Additional main-chain NH groups at positions 4 and 5 form hydrogen-bonds to the coordinating oxygen atom. In light of the extensive hydrogen bonding to both oxygen atoms of this aspartate side chain, it is clear to see that even the conservative change to an asparagine would not be tolerated. It would result in considerable reorganization of the loop with concommitant loss of Ca^{2+} binding affinity.

Position 2 ($<\phi,\psi>$ = -65°, -37°).

The nature of the residue in this position is not highly conserved. However, in approximately 50% of the loops, it is a basic residue, a lysine or less frequently an arginine (Table 3.2). The side chain of lysine 143 on loop IV of TnC extends into the solvent (Figure 3.5) and is directed towards the C-terminus of the incoming helix so that the positive charge on this basic residue has a strong stabilizing effect upon the helix dipole. It is curious that in loop II of the TnC molecules from different species this residue is a conserved glutamate, a negatively charged side chain that should have a pronounced destabilizing effect on the C helix. In fact, in turkey TnC there is no electron density for the side chain of Glu67, indicating that this residue is disordered and highly mobile. In the C-terminal domain of TnC (the high affinity Ca^{2+}-binding sites) both loops have a basic residue in position 2. In addition to the γ-turn hydrogen bonding, the NH of this residue forms a bifurcated H-bond with the carboxylate oxygen of the glutamate in position 12 (Figure 3.5).

Position 3 ($<\phi,\psi>$ = -98°, 7°; $<\chi_1,\chi_2>$ = 67°, 11°).

The residue in position three (Y) is most commonly an aspartate or an asparagine. It forms a direct coordination to the Ca^{2+} ion and contributes to the pentagonal plane. The main chain nitrogen atom forms a relatively long hydrogen bond to the carboxylate of the invariant glutamate at position 12 (Figure 3.5). In those cases where the residue at

position 3 is an asparagine, the side chain amide can form a hydrogen bond to either the water molecule in the -X position or to the side chain of the coordinating residue in position 5. This latter interaction between Asn144 $N^{\delta 2}$ and Asp146 $O^{\delta 2}$ of loop IV of turkey TnC can be seen in Figure 3.5.

Position 4 ($<\phi,\psi>$ = 58°, 36°).

The most common residue in this position is a glycine. This is commensurate with the ϕ,ψ angles. Other residues that are found in position four include asparagine, alanine and lysine. A bifurcated hydrogen bond from the main-chain NH is made to the carbonyl oxygen atom of the aspartate in position 1 (a Type I 3_{10} turn) and to the Ca^{2+}-coordinating oxygen atom of the same aspartate. Both distances are long, ~ 3.3 Å on average for the several loops analyzed. In loop two of CaM the distances are especially long but the directions are still such that a favorable electrostatic interaction could result.

Position 5 ($<\phi,\psi>$ = -90°, -3°; $<\chi_1,\chi_2>$ = 65°, 10°).

Position 5 is also a Ca^{2+}-binding ligand (Z in Table 3.2). The common residue is an aspartate or asparagine, but serine and glycine have also been observed. Serine at this position is particularly interesting. In the parvalbumins almost all Ca^{2+} binding loops I have a serine at this position. Not only does the O^{γ} atom form a coordinating ligand to the Ca^{2+} ion, but also it forms a strong hydrogen bond to the carboxylate group of the glutamate in position 9 (distance $O^{\gamma}\cdots O^{\varepsilon 2}$ 2.8 Å, Figure 3.4c). A glutamate is the only amino acid with a sufficiently long side chain to coordinate directly to the Ca^{2+} from position 9 (157). For those Ca^{2+} ions that have a water molecule in the -X position, the side chain of the residue in position 5 is favorably disposed to accept (aspartate) or donate (asparagine) a hydrogen bond with the water. The main chain NH of residue 5 forms, on average, a bifurcated hydrogen bond with the oxygen atoms that coordinate to the Ca^{2+} ion of residues in positions 1 and 3 (Figure 3.5). The interaction with the side chain of position 3 is an n to n+2 type (156). For those loops in which adjacent positions 1 and 3 both contain a charged aspartate, this favorable electrostatic interaction is highly stabilizing.

Position 6 ($<\phi,\psi> = 90°, 2°$).

This position is occupied by an invariant glycine. The main chain NH forms the important hydrogen bond to the non-coordinating oxygen atom of the carboxylate of aspartate in position 1 (Figure 3.5). This is a relatively strong hydrogen bond with an average N to O distance of 2.7 Å in the 11 loops. The main chain conformation for position 6 is a common conformation for a glycine residue. It assists the chain to make a 90° turn in direction so that the remaining Ca^{2+} ligands are in coordinating positions. The carbonyl-oxygen atom of glycine 6 forms a hydrogen bond with a conserved water molecule in 7 of the 11 loops. This water, in those structures for which it was selected as a well ordered solvent, forms a hydrogen-bonded bridge to the adjacent loop of the pair at the main chain NH of the residue in position 10, the first residue of the exiting helix. It is not present in the Ca^{2+}-free loops of the avian TnC molecules.

Position 7 ($<\phi,\psi> = -126°, 157°$).

From the ϕ,ψ values it is apparent that the residue in this position initiates the small β-strand that extends for three residues. Of all positions in the calcium binding loop this is the most variable in terms of the nature of the amino-acid. The side chains extend into the solvent. However, there does seem to be a fairly common electrostatic interaction between aromatic groups (phenylalanine or tyrosine) in the first loop of a high affinity pair and a basic residue (arginine or lysine) in the second loop. This interaction is typified by loops III and IV of TnC and loops I and II of Parv (Figure 3.4b,c). Loops III and IV of CaM have a tyrosine and glutamine respectively that are similarly disposed (Figure 3.4e).

Position 7 also provides its main chain carbonyl oxygen atom to the coordination sphere of the Ca^{2+} ion at -Y (see Table 3.2 and Figure 3.4). The β conformation allows both of the NH and the CO groups of this residue to point in towards the centre of the loop. The NH forms an n to n-2 hydrogen bond with the Ca^{2+}-coordinating oxygen atom in the side chain of the residue in position 5. This hydrogen bonding interaction, along with the others mentioned above, serves to stabilize partially the close proximity of negatively

charged oxygen atoms in the coordination sphere of the calcium.

Position 8 ($\langle\phi, \psi\rangle$ = -108°, 120°).

Position 8 is the central residue of the short β-strand. In all of the loops this residue has a hydrophobic character with a preponderance for isoleucine. The side chains of the residues in position 8 penetrate the protein core and seem to form a central point around which the loops can adjust their conformations (158,159). The main chain NH and CO groups of the residue in position 8 face away from the central cavity of the loop towards the neighboring loop. Antiparallel β-sheet type interactions are formed with the main chain amide and carbonyl group of the adjacent strand also at position 8. The pseudo two-fold axis relating the two loops in the pair of EF-hands passes approximately between these two residues at positions 8. The cooperativity in Ca^{2+}-binding (101,160,161) may involve this region of the loop and likely includes the residue at position 7.

Position 9 ($\langle\phi,\psi\rangle$ = -94°, 171°; $\langle\chi_1,\chi_2\rangle$ = 68°, 16°).

It appears that the residue in position 9 has several functions relative to the stability of the loop in a Ca^{2+}-binding conformation. It was originally proposed as one of the coordinating ligands to Ca^{2+} (29) and indeed in the parvalbumin loop I, the glutamate at this position (-X) does perform such a role (Figure 3.4c). On the other hand, in the majority of cases when the side chain at position 9 is a Ser, Thr, Asp or Asn, its side chain is too short to coordinate directly to the Ca^{2+} and the ligand at the -X position is always a water molecule (157). This is even true in the case of the less common variant loop I of ICaBP (Figure 3.4f). The side chain of the residue in position 9 does not always form a hydrogen bond to the coordinating water molecule. However, it does seem to be the residue that initiates the exiting helix. This initiation (or stabilization) involves the side chain of the residue in position 9 forming a hydrogen bond to the NH of the invariant glutamate in position 12 [an n to n+3 interaction detailed by Baker and Hubbard (156)]. Several of the loops also have n to n+2 hydrogen bonds from the position 9 side chain to the NH of the residue at position 11, but the most common interaction is the n to n+3

(Figures 3.4 and 3.5). The majority of helices in the CaM family of proteins are initiated by similar side chain (Ser, Thr, Asp or Asn) to main chain NH hydrogen bonding interactions (see Table 3.2). Parvalbumin is a notable exception with Glu59 and Gly98 in the initiating positions.

Another very important interaction is provided by the main chain of position 9. The NH forms a hydrogen bond to the carboxylate group of the invariant glutamate (Figure 3.5). Thus, in the Ca^{2+}-bound form of these loops, the two invariant residues, aspartate at position 1 and glutamate at position 12, are the recipients of many stabilizing, favorable electrostatic interactions. The carbonyl oxygen of the residue in position 9 is involved with a normal 1-5 α-helix type hydrogen bond in the exiting helix.

Position 10 ($\langle\phi,\psi\rangle$ = -60°, -41°).

The residue in this position is the first one of the exiting helix with helical ϕ,ψ values and normal α-helical 1-5 hydrogen bonding for its carbonyl oxygen atom. The amino nitrogen of this residue is often hydrogen bonded to a conserved water molecule (Figure 3.5). There is a preponderance of aromatic residues at position 10 in loop II or loop IV of a pair of EF hands (Table 3.2, Figure 3.4). In these cases, the aromatic side chain forms part of the aromatic cluster that involves the two aromatic residues at the C-terminus of the incoming helix (at -1 and -4) of loops I and III and the two aromatics of the exiting helix of loops II and IV (position 10 of the loop and the residue following the invariant glutamate (Table 3.2).

Position 11 ($\langle\phi,\psi\rangle$ = -63°, -44°).

The residue in position 11 is most commonly a negatively charged residue, an aspartate or glutamate, although in some loops it is a basic residue, lysine in loop I of TnC and CaM. The negatively charged carboxylate groups are electrostatically stabilizing for the amino-terminus of the exiting helix. Often the NH of this residue forms a hydrogen bond to a water molecule (Figure 3.5).

Position 12 ($\langle\phi,\psi\rangle$ = -66°, -41°; $\langle\chi_1,\chi_2,\chi_3\rangle$ = -72°, 166°, -25°).

The last position in the Ca^{2+}-binding loop is also the third position of the exiting helix. Position 12 is the invariant glutamate that has both oxygen atoms of its carboxylate group coordinating to the Ca^{2+} ion in a bidentate manner (Figure 3.3b and Figure 3.4a-f). As described above this residue could be the key residue for adapting the metal-binding loop towards the binding of either Mg^{2+} or Ca^{2+} ions. The long range hydrogen bonds to the carboxylate oxygen atoms from main-chain NH groups of the residues in positions 2, 3 and 9 have already been described.

Until recently, it was thought that the Ca^{2+}-binding loop conformation described above only occurred in the HLH family. However, the crystal structure of the galactose binding protein revealed a very similar tertiary structure for the Ca^{2+}-binding loop (28) in spite of the different secondary structures of the flanking segments of polypeptide chain. The loop consists of nine residues (Asp134 to Gln142), each alternate residue contributing a ligand to the Ca^{2+} ion. The final two ligands come from both oxygen atoms of the carboxylate of a neighboring glutamate (Glu205). The seven oxygen ligands are arranged in an approximate pentagonal bipyramidal fashion analogously to the one found in EF hand loops. A least-squares overlap of the 44 common main chain atoms in the galactose binding protein with the main chain atoms of the EF site of parvalbumin yielded an rms deviation of 0.60 Å, indicating the remarkable structural equivalence (28). Hydrogen-bonding interactions within the loop are similar to those shown in Figure 3.5 for loop IV of TnC. The similarity extends even to the stabilizing main chain hydrogen bonds from Asn136NH and Gln142NH to both oxygens of the Glu205 carboxylate group. Although not explicitly stated in the paper, it appears that the isoleucine at position 8 in the galactose binding protein loop forms β-type hydrogen-bonding interactions with an adjacent strand. However, this β strand is parallel and not antiparallel as in the pair of loops in the HLH motif proteins. The branched aliphatic side chain contributes to the hydrophobic interior of the domain, as in the other proteins.

N-Terminal Loops of TnC — The N-terminal loops of TnC are metal free in both

the turkey and chicken crystal structures (41,43). As a result, they have quite a different conformation to the loops with Ca^{2+} bound (Figure 3.4a) (157). The rms difference between the N- and C-terminal loops of TnC (96 main-chain atoms in the comparison) is 3.07 Å (Table 3.5). On the other hand, if one compares individual loops (i.e., I with III and IV and II with III and IV) the rms deviations (48 atoms compared) are ~ 1.9 Å (43,157). Furthermore, if one limits the comparisons to the first six residues of the loops or the last six residues of the loops, then rms deviations of ~ 0.6 Å result. This suggests that the Ca^{2+} ligands in each half of the loop are close to their optimal position for Ca^{2+} binding and only need to come together for it to take place (40,157,162). However, it is not a simple hinge motion; rather, cumulative small differences in ϕ and ψ over several residues in the central portion of the loop bring the two parts of the loop together to form the completed site. These Ca^{2+}-dependent differences in conformation of the loop would have attendant changes in the interhelical angles of the whole domain when Ca^{2+} ions bind.

In addition to the two antiparallel β-sheet hydrogen-bonding interactions at position 8 of the Ca^{2+}-filled loops, the Ca^{2+}-free loops I and II have a third hydrogen bond. This bond, from the NH of the residue at position 10 direct to the CO of the invariant glycine in the adjacent strand, extends the β-sheet. In the Ca^{2+}-filled conformation, the equivalent residues at positions 10 and 6 are also hydrogen-bonded but through a conserved water molecule (Figure 3.5).

The Variant Loop of ICaBP — ICaBP has the common HLH motif for binding site II but the loop of binding site I departs from the normal structure (33,34). This structural difference has also been inferred from solution NMR studies (163). In spite of the difference, this loop also has seven ligands that coordinate to the Ca^{2+} ion. Table 3.2 shows that there is no sequence homology with the other EF hand structures for the first part of loop I. The Ca^{2+}-binding ligands at X, Y and Z do not come from side chain oxygen atoms as in the common EF hands; the ligands are carbonyl oxygen atoms of the main chain at Ala14, Glu17 and Asp19. Therefore the negative charges arising from

aspartate residues in the coordination sphere of Ca^{2+} are replaced by peptide dipoles with the negative pole directed towards the Ca^{2+} ion. There are two insertions in the first part of the loop such that the right-angled bend in the loop normally provided by the invariant glycine at position six is now at the eighth amino acid of the loop (Asn21). In addition, the conformational angles (ϕ,ψ) of this asparagine are not the same as those of the invariant glycine.

There are stabilizing hydrogen bonds for the first part of the loop (Figure 3.4f). The equivalent residue to the invariant glutamate (Glu27) is the recipient of a favorable electrostatic interaction from the main chain NH of Glu17 (position Y) and a hydrogen bond from the NH of Ser24, the equivalent to position 9 in the commonly observed loop. The side chain of Asp19, which does not provide the ligand, is involved with an n to n+2 hydrogen bond to the main chain NH of Asn21 and in a hydrogen bond to the side chain of Gln22 (equivalent to position 7). There is a long favorable interaction from the NH of Gln22 to the CO of Asp19 in the Z position; this has an analogous interaction in the previously described structures.

The last six residues of loop I in ICaBP (Gln22 to Glu27) are structurally similar to those equivalent residues in the standard loops. The -X position of the Ca^{2+}-coordination sphere is also a water molecule. The only differences are the side chain interactions of the residues in position 7 of loop I and loop II. Each one (Gln22 and Glu60) reaches across to the adjacent loop (Figure 3.4f) and forms a hydrogen bond to the water molecule coordinating the Ca^{2+} ion (-X). In spite of the very different tertiary structure that loop I has from the standard loop, the pK_d for Ca^{2+} is very similar to that of loop II and to other EF hand loops, ~ 6.6 (163,164).

HELICES

Helix Regularity — Figures 3.2a-d clearly show that the predominant secondary structure in the Ca^{2+} modulated proteins is α- or 3_{10}-helix. Analysis of the main-chain dihedral angles and hydrogen bonding patterns indicate the following percentages of

residues in helical conformations: TnC, 68%; CaM, 66%; Parv, 56% and ICaBP, 59%. For the most part the helices are amphipathic and display the regular n to n+4 hydrogen bonding pattern typical of α-helices. However, there are notable exceptions in various helices of all four proteins. The central helix of CaM is irregular and curved. The B helix of TnC, the D helix of parvalbumin, and the D helix of ICaBP are markedly kinked. The latter three forego their α-helical conformations in specific regions in order to maximize hydrophobic interactions with adjacent helices.

Figure 3.6a shows a comparison of the conformations of the central parts of the two helices that connect the N- and C-terminal domains of TnC and CaM. These long helices are comprised of the relatively hydrophobic D helix, the highly charged D/E linker and the amphipathic E helix; in total, continuous helical segments of 31 and 28 residues, respectively. In TnC, the entire length of the helix is relatively straight [a slight bend with a radius of curvature of 137 Å (41)]. Average values of ϕ and ψ are -60° and -45° and the hydrogen bonding is the normal n to n+4 of an α-helix. The only aberration in α-helical parameters occurs at the central Lys91-Gly92-Lys93 triplet where ϕ, ψ values are -47° and -56° (Figure 3.6a). Also in this region, Ser94 O^{γ} provides an additional hydrogen bond to the main-chain carbonyl oxygen of Lys91.

The central helix of CaM is much less regular in its conformational angles (38). There are two places in which the helical parameters are disrupted, at Thr70 and at Thr79. The environment of the second threonine is visible in Figure 3.6a. For both of them the n to n+4 hydrogen bond is long (3.53 Å) and the ϕ, ψ values are atypical of α-helical parameters, Thr70 (-74°, -34°), Thr79 (-100°, -12°). In the region from Lys77 to Glu84 there are several bifurcated hydrogen bonds. These residues are therefore in a mixed conformation between 4_{13} and 3_{10} helices. There are also some unambiguous 3_{10} turns with n to n+3 hydrogen bonding. As a result, the carbonyl oxygen atoms in this region are splayed out towards solvent (Asp78 O, Thr79 O, Asp80 O). The side chain O^{γ} atom of Ser81 forms a hydrogen bond to Lys77 O; the main chain has conformational angles ϕ, ψ

of -93°, -63° and its carbonyl oxygen, Ser81 O, accepts two hydrogen bonds, one from Ile85 N (2.99 Å) and the other from Glu84 N (2.82 Å) (see Figure 3.6a).

The residues of the central long helix in both CaM and TnC exhibit some of the highest temperature factors in these molecules (38,41,43). These crystallographic studies suggest conformational flexibility in the central helix. Neither in TnC nor in CaM does this segment of the helix have close intermolecular packing contacts in the crystals that might warrant its distortions.

Some biochemical and biophysical studies have also suggested flexibility in the central helices of TnC and CaM (145,146,165,166). Circumstantial support for flexibility of the central helix comes from the site of proteolytic cleavage in both CaM and TnC. The major sites of hydrolysis by bovine trypsin in these two molecules are at Lys77 and Lys87, respectively (167,168). The conformational specificity of trypsin for substrates would require that the central helices of TnC and CaM would partly unfold in order for them to fit into its active site. In addition, one study has shown that intramolecular crosslinking of the N- and C-terminal domains of CaM could be achieved by replacing Gln3 and Thr146 with cysteine residues (169). The authors suggest that the major bend in the central helix required for the crosslinking could occur by adjusting the backbone ϕ, ψ angles of Thr79. Furthermore, they have used the concept of helix bending to formulate a model for the interaction of CaM with its target enzymes (170). However, the issue of central helix bending in the modulation of TnC and CaM activity is still very controversial (1,51). Two recent site specific mutagenesis studies of the central helices on TnC and CaM have provided interpretations that are contrary to helix bending in the function of these proteins. Replacement of Thr79 in CaM and Gly92 in TnC by proline residues did not significantly alter the activities of these proteins towards their target molecules (123,143).

The B helix of TnC has a pronounced bend (Figure 3.6b) that is accommodated by the insertion of two solvent molecules into the helix path (43). Helix B is the exiting helix from loop I in TnC. Normally the exiting helices are initiated by the side chain of the

residue in position 9 of the loop (see above). In the B helix however, Ser38 O^{γ} forms an n to n+2 hydrogen bond with Lys40 N rather than with the NH of Glu41, as it would when initiating a helix. In addition, the NH of Glu41 forms an n to n+3 hydrogen bond to the carbonyl oxygen of Ser38, in a type III β-turn. This is followed by a type I β-turn in which the NH of Leu42 forms an n to n+3 hydrogen bond to the CO of Thr39. It is at this point that the bend occurs. Solvents O181 and O183 assume the structural positions of the carbonyl oxygen atoms of Thr39 and Lys40 in an undistorted helix. These solvent atoms are involved in extensive hydrogen bonding interactions (Figure 3.6b) that presumably stabilize the irregular helix. From Gly 43 on, the remaining residues of helix B adopt a regular α-helical structure. A major consequence of the bend in helix B is to augment the hydrophobic interactions made by Leu42, Val45 and Met46 with the hydrophobic surfaces of helices A and D. Only in the Ca^{2+}-free form are these three helices in sufficiently close proximity to form the hydrophobic interactions. Therefore the lack of Ca^{2+} binding in loop I may be directly related to this helix distortion.

Helix D in Parv is decidedly kinked with a bend angle of 124° (Figure 3.6c). As for the B helix of TnC, an ordered water molecule is found in a compensating position where distortion of the helix precludes normal main-chain hydrogen bonds. Glu59, because of its involvement with Ca^{2+} coordination in loop I, does not initiate the helix. From Glu60 to Lys64, the main-chain conformation is essentially α-helical. There is a type III β-turn formed by residues Glu61 to Leu65. This is followed by a type I β-turn with the NH of Phe66 forming a hydrogen bond to the CO of Leu63. The distortion in the path of the polypeptide chain occurs at Leu65 (ϕ, ψ; -97°, 13°). The water molecule (O118) bridges the main chain NH of Leu67 and the carbonyl oxygen atom of Lys64 that would normally be directly hydrogen-bonded in 3_{10} helical conformation (Figure 3.6c). The remaining residues of the D-helix adopt a 3_{10} helical conformation. Water O118 also forms a hydrogen bond with the main chain carbonyl oxygen atom of Arg75.

Enhanced hydrophobic interactions are facilitated by the bending of helix D.

Phe66, Leu67 and Phe70 form a hydrophobic array with which the four phenylalanine residues on the A/B domain interact.

Helix D of ICaBP begins as a regular α-helix that is initiated normally by the hydrogen bond from Oγ of Ser62 to the NH of Glu65 (34). At residue Phe66, the helix bends to an angle of 150°, adopts a 3_{10} conformation for one turn, makes a further bend to 130° and finishes with the final 6 residues in an α-helical conformation. The bends allow helix D to pack tightly against the rest of the molecule and close in over the hydrophobic surface. As in the other cases, the favorable hydrophobic contacts formed presumably stabilize the distorted helix.

The above description of these several distorted helices suggests that α-helices are highly adaptable to specific environments. Large distortions in which the helix bends by up to 60° (from 180°) can be achieved by altering the main chain conformational angles of the involved residues by only a few degrees, with the concomitant change of the hydrogen bonding interactions from a 4_{13} to a 3_{10} pattern. Solvent molecules such as water are ideally suited to compensate for the loss of main-chain hydrogen-bonding that result from such rearrangements.

Helix Packing — The four helices of the Ca^{2+}-binding domains pack against one another in a highly conserved fashion (Table 3.6, Figure 3.1). Adjacent helices within a single HLH motif, i.e. E/F or G/H, are approximately perpendicular (range, 84° to 109°). They contact each other only at the ends, where they are joined to the Ca^{2+} binding loop (Figures 3.1, 3.2a-d). Adjacent helices from two separate HLH motifs (F/G and E/H) pack together with an average crossing angle of 118° (range, 110° to 124°). The main interhelix axis contact distance for these helical pairs is 9.9 Å (Table 3.6). The overall outcome of the above packing scheme for the Ca^{2+}-bound domains is that there is a much larger, more favorable interaction area between helices from neighboring HLH motifs (E/H and F/G) than for those within a single HLH motif.

The large variation in the interhelix axis crossing angles between E/F or G/H pairs

suggests major conformational differences in the Ca^{2+}-binding domains. However, the data in Table 3.4 show that this is not the case and the Ca^{2+}-bound domains are very similar. The exception is ICaBP (34) which has much larger interhelical angles within HLH motifs (129° and 119°). These angles are a result of several key hydrophobic contacts between helices B, C and their linking peptide with helix D that curves in towards them (Figure 3.2d). ICaBP is the only Ca^{2+} bound domain in which all four helices interact with one another.

The interhelical angles between helices on adjacent HLH motifs (A/D and B/C) of the Ca^{2+}-free N-terminal domain of TnC maintain the average packing angle of 118° seen in the Ca^{2+}-bound domains (48,162) (Table 3.6). On the other hand, the crossing angles between helices within a single HLH motif are much larger than those of the Ca^{2+}-filled domains.

Interhelical interactions are much more favorable in the Ca^{2+}-free form. Adjacent helices within an HLH motif become nearly antiparallel instead of roughly perpendicular. This increases energetically favorable interactions in several ways. The contact area amongst all four helices in the N-terminal domain increases relative to that of the corresponding Ca^{2+}-bound domain. The helix dipoles are more antiparallel and therefore provide favorable electrostatic interactions along their length. In the Ca^{2+}-filled form, adjacent helices cross only at their ends; therefore, helix dipole interactions are limited. Furthermore, interhelix hydrophobic interactions are maximized and there are no exposed hydrophobic surfaces as in the C-terminal domain. Thus, in the absence of the favorable forces associated with Ca^{2+} binding, the helices seem to adopt a lower energy conformation with increased stabilizing helix-helix interactions (171,172).

HYDROPHOBIC INTERACTIONS

All four proteins have a hydrophobic core located at the interface of the two HLH motifs.[1] The two conserved hydrophobic residues in position 8 of the adjacent Ca^{2+}-

[1] In the following discussion, we refer to the four helices of the Ca^{2+}-binding domains in

binding loops form the nucleus of this core (Table 3.2, Figures 3.7a-f). These two residues interact not only with each other to stabilize the loops, but also form several key contacts with conserved hydrophobic residues in each of the four helices.

The interface between helices E and H consists of a conserved set of hydrophobic interactions that involve adjacent phenylalanine residues (Figures 3.7b-f). The aromatic side chains of these residues (Phe102 and Phe154) pack against one another in the same manner in all four proteins. In both domains of CaM and TnC and in ICaBP, a third aromatic group (Phe151) stacks against the first two. The characteristic aromatic ring stacking, edge to face and roughly perpendicular, is observed in the cluster. In Parv, the analogous position is occupied by Val99 which also interacts with the adjacent phenylalanine side chains (Figure 3.7d).

Whereas the E/H interface is predominantly aromatic, the interface between helices F and G is aliphatic in nature (Figures 3.7b-f). The consistently larger interhelical distances between helices E and H (Table 3.6) may be a reflection of this phenomenon. Unlike the extensive and highly conserved packing between the E and H helices, the hydrophobic interactions between the F and G helices are more variable and often involve the methylene carbon atoms of charged residues. The conserved hydrophobic interactions between the HLH motifs extending from the E/H helices, past the loop and to the F/G helices, form the structural interior of all four molecules regardless of variation in primary sequence, interhelical angles or distortions in secondary structure. Some of these contacts had been suggested earlier based on solution studies using NMR methods (174).

In addition to the hydrophobic core, the C-terminal domain of TnC and both domains of CaM have large expanses of hydrophobic surface exposed to solvent (38,48, 162). Some of the residues in these regions have been implicated in the binding of target molecules (38,61,92,94,142,175-180). The homologous residues of the Ca^{2+}-free N-

---

the generic sense as shown in Figure 3.7a (E,F,G,H). For convenience, the nomenclature given in Figure 3.7a is for the C-terminal domain of TnC; the equivalent positions for the other domains can be gleaned from Table 3.2.

terminal domain of TnC, of Parv and of ICaBP are also hydrophobic but not accessible to solvent.

Figures 3.7b and 7c display the hydrophobic surfaces of the C-terminal domains of TnC and CaM. The sequence (Table 3.2) and structural homologies of these surfaces is evident. There are many methionine residues in this region, four in each of the domains of CaM and two in the C-terminal domain of TnC. NMR studies have shown that the chemical shifts of methionine and phenylalanine residues are especially sensitive to complexation of CaM with the CaM-binding domain of MLCK (177) and with trifluoperazine, a potent CaM antagonist (175). Furthermore, chemical modification of several methionine residues in the N-terminal domain of CaM (e.g. Met71, Met72 and Met76) inhibited its activation of cyclic nucleotide phosphodiesterase (178,146), adenylate cyclase (179), erythrocyte Ca^{2+}-ATPase (176), as well as inhibiting the binding of trifluoperazine (180). This latter study also showed that nitration of Tyr99 and Tyr138 (Figure 3.4e), which are on the totally opposite surface of the molecule to the hydrophobic patch (38,142), has no effect on the ability of CaM to activate phosphodiesterase.

Conserved hydrophobic interactions involving the surface residues of TnC and CaM are shown in Figure 3.7a. A residue involved in one of these contacts, Phe105 of turkey TnC (Figure 3.7b), exhibits a major difference in conformation from the homologous residues in CaM (Figure 3.7c) and chicken TnC. In turkey TnC this residue has a high energy conformation with $\chi_2 \approx 5°$. In each of the latter two proteins, this phenylalanine adopts the minimum energy conformation with $\chi_2 \approx 90°$. It appears that chicken and turkey TnC have sampled conformational space in a different manner; neighboring residues Cys101 and Ile121 also adopt alternative conformations. In turkey TnC the mean B-factors for Phe105 are ~ 10 $Å^2$ (41); in chicken, the equivalent atoms have B-factors of 40 $Å^2$ (43).

In the Ca^{2+}-free domain of TnC, homologous hydrophobic residues are present (Figure 3.7f, Table 3.2). The N-terminal helix (N) is stabilized by the presence of several

additional hydrophobic residues on helices A (Met18, Ile19) and D (Val80 and Val83). There are significant differences in the spatial distributions of some homologous hydrophobic residues due to the changed disposition of the helices in the N-terminal domain (cf., Figures 3.7b and 3.7f). As a result, this domain exhibits many more interhelical hydrophobic interactions and no large exposed hydrophobic surfaces.

Like the N-terminal domain of TnC, Parv does not have a large exposed hydrophobic surface. In the case of Parv, this is not the result of changed helix disposition, but is due to the folding of the additional 39 residues of the A/B domain to fill the hydrophobic cup formed by helices C,D,E and F of the Ca^{2+} binding domain (Figure 3.2c, Figure 3.7d). The hydrophobic residues of helices A and B are accommodated by residues lining the hydrophobic cup unique to the parvalbumins (Table 3.2). Phe66, 70, 85 and Leu77 interact with Phe24, 29 and 30 in the classical herringbone fashion. Perhaps the presence of the N-terminal 39 amino acids in Parv aids in stabilizing the Ca^{2+}-bound conformation of the domain by burying the hydrophobic surface.

Like the N-terminal domain of TnC, the disposition of helices in ICaBP buries the potential hydrophobic surface (34). The distortion of the D-helix along with the hydrophobic residues on the extended peptide linker (Phe36, Leu39 and Leu40) contribute to the stabilization of the more "closed" conformation (Table 3.6, Figure 3.7e).

ELECTROSTATIC INTERACTIONS

The charged residues in the four HLH Ca^{2+}-binding proteins form electrostatic interactions with the metal ions, the bulk solvent or, less frequently, with one another. Negatively charged carboxylate groups provide many of the oxygen ligands for the Ca^{2+} ions. In addition, there are two other regions in which they are clustered together. Both TnC and CaM exhibit these clusters on the N-terminal portions of helices A, C, E and G (Table 3.2). These negatively charged patches surround the hydrophobic surfaces discussed in the previous section (Figures 3.7a-f). The conserved nature of these charged regions in TnC and CaM suggest that they may play an important role, along with the

hydrophobic surfaces, in interacting with target molecules. Indeed, the proposed TnC binding site on TnI (51,60), the CaM binding site on MLCK (181), the CaM binding site on phosphofructokinase (182) and several of the antagonists of TnC and CaM (183,184) exhibit clusters of positive charge complementary to the negatively charged clusters on TnC and CaM. Several computer studies that have modelled the interactions of target molecules and drugs to CaM have also implicated the negatively charged regions (142,185-187). More direct evidence of the importance of the charge cluster on helix E in CaM comes from site directed mutagenesis in which the three glutamates were changed to lysine residues (188). The resulting mutant CaM was unable to activate the myosin light chain kinase effectively.

Positively charged amino-acids do not form specific clusters as do the negatively charged groups. The most conserved lysine is in position 2 of the loops (vide infra). In each of the domains of CaM there is a positively charged NH_3^+ group from Lys75 and from Lys148 that extends over the hydrophobic surface. Differential labelling of these lysine residues was important in pinpointing the sites of interactions of the target enzymes with Ca^{2+}-calmodulin (189-193).

The contribution of electrostatic interactions among charged side chains to protein structure and function is difficult to evaluate. A complete description of the solvation state, the presence of metal ions, the net charge of the group and the mobilities are needed in conjunction with the results of the crystal structure determinations in order to define what constitutes a significant charge-charge interaction. Some of these problems have been addressed in other systems (194,195). Until similar calculations will have been done with the HLH family of Ca^{2+}-binding proteins, a meaningful description of the ion pairs is of necessity incomplete.

Examination of the structurally aligned sequences in Table 3.2 shows many conserved sites for the charged residues. This implies that there would also be a number of conserved ion pair interactions among them. In general, though, this is not the case. For

example, the ion pair interaction between Arg84 and Glu64 is present in both avian TnC molecules. In CaM, the side chains of the homologous residues are 10 Å apart and in ICaBP, Lys72 and Glu52 interact through a bridging water molecule. Parv has neither of these residues.

The sequences also show the presence of oppositely charged amino acids two or three residues apart in some of the helices. It has been suggested that these salt bridges provide stabilization of the long central helices of TnC and CaM (196). Only two of the seven originally predicted are confirmed as direct hydrogen-bonded ion pairs in the refined TnC structures (43,41). The others, although too far apart to be considered direct salt bridges, could still provide favorable electrostatic interactions and enhance the level of ordered solvent in this helical region (197).

CONFORMATIONAL CHANGE OF TNC

There is ample evidence that TnC undergoes a conformational change when it binds Ca^{2+} ions (50-52). A model for this conformational change has been proposed (48). It was based upon the extensive structural similarity of the Ca^{2+}-filled domains (Table 3.4) and the high degree of sequence homology between these domains and the Ca^{2+}-free N-terminal domain of TnC (Table 3.2). In that proposal, the loops and helices of the N-terminal domain would adopt the disposition of those of the C-terminal domain upon Ca^{2+}-binding (48,162).

Figures 3.8 and 3.9 show the essence of this proposal. The Ca^{2+}-free conformation, represented by the N-terminal domain of the TnC crystal structures, has a closed conformation, in which the hydrophobic residues of the surface are inaccessible to solvent (Figure 3.8a). Upon Ca^{2+}-binding, the N-terminal domain would adopt a more open conformation (Figure 3.8b). The transition requires a reorientation of the helices within a single HLH domain (A/B, C/D). Residues that comprise the linker peptide between helix B and helix C move by up to 14 Å (48,162). Adjacent helical pairs A/D and B/C retain their relative orientations and interhelix crossing angles (Table 3.6). This is a reflection of the

more extensive hydrophobic interactions that bind these pairs together. The proposal also specifies that helix N does not change its orientation relative to helix D. Such a model for the conformational change involves only a rearrangement of existing secondary structural units.

The model also proposes that Ca^{2+}-binding to the N-terminal domain would not have much effect on the conformation of the C-terminal domain. This is consistent with the results from numerous circular dichroism and ^1H NMR studies done on whole TnC and on each of the isolated N- and C-terminal domains (50,158,198). Additionally, the similarity of the overall dimensions of the proposed 4-Ca^{2+}-model with the 2-Ca^{2+} crystal structures is strongly supported by the recent solution X-ray scattering data (146).

The opening of the N-terminal domain upon Ca^{2+} binding exposes a prominent hydrophobic surface analogous to those of the TnC C-terminal domain and both domains of CaM (Figures 3.9 and 3.7b and c, respectively). Exposing these hydrophobic surfaces is attended by large changes in the environments of Phe22, Phe29, Val45, Leu49, and Glu57 (Figures 3.9a,b). These movements are supported by an earlier NMR study on Ca^{2+} titration of the N-terminal sites of TnC that showed large changes in the chemical shifts of valine, leucine, and two or more phenylalanine and glutamate residues (63). Figure 3.9 also shows that the side chains of Met82 and Glu85 on helix D, and of Val45, Met46, Met48 and Leu49 on helix B have dramatically increased solvent accessibilities (48,162). These residues, as well as Met18, Phe29 and Met86, constitute the exposed hydrophobic surface that, along with the negatively charged clusters of Glu16, Glu21, Glu56, Glu57 and Asp59 are implicated in Ca^{2+}-mediated target molecule binding.

ACKNOWLEDGEMENTS

We would like to thank Drs. Swain, Amma and Kretsinger for supplying us with atomic coordinates of the refined carp Parv prior to their public release; Drs. Babu, Cook and Bugg for the use of the refined coordinates of bovine brain CaM and a preprint of their paper on the refined structure. Tony Hawreletchko, Mark Israel and Perry D'obrenan

helped with the preparation of various figures. Mae Wylie handled the many revisions of this manuscript with her usual expertise. NCJS acknowledges the support of AHFMR for a studentship. This research was supported by the Medical Research Council of Canada.

REFERENCES

1. Forsén, S. 1988. In *Inorganic Biochemistry*, ed. I. Bertini, Univ. Sci. Books, Mill Valley, CA. In press
2. Bode, W., Schwager, P. 1975. *J. Mol. Biol.* 98:693-717
3. Bier, M., Nord, F. F. 1951. *Arch. Biochem. Biophys.* 33:320-332
4. Cliffe, S. G. R., Grant, D. A. W. 1981. *Biochem. J.* 193:655-658
5. Read, R. J., James, M. N. G. 1988. *J. Mol. Biol.* 200:523-551
6. Russin, D. J., Floyd, B. F., Toomey, T. P., Brady, A. H., Awad, W. M. Jr. 1974. 249:6144-6148
7. Olafson, R. W., Smillie, L. B. 1975.*Biochemistry* 14:1161-1167
8. McPhalen, C. A., Svendsen, I., Jonassen, I., James, M. N. G. 1985. *Proc. Natl. Acad. Sci. USA* 82:7242-7246
9. McPhalen, C. A., Schnebli, H. P., James, M. N. G. 1985. *FEBS Letters* 188:55-58
10. McPhalen, C. A., James, M. N. G. 1988. *Biochemistry* 27:6582-6598
11. Bode, W., Papamokos, E., Musil, D. 1987. *Eur. J. Biochem.* 166:673-692
12. Matsubara, H., Hagihara, B., Nakai, M., Komaki, T., Yonetani, T., Okunuki, K. 1958. *J. Biochem.* 45:251-255
13. Matthews, B. W., Weaver, L. H., Kester, W. R. 1974. *J. Biol. Chem.* 249:8030-8044
14. Holmes, M. A., Matthews, B. W. 1982. *J. Mol. Biol.* 160:623-639
15. Feder, J., Garrett, L. R., Wildi, B. S. 1971. *Biochemistry* 10:4552-4556
16. Voordouw, G., Roche, R. S. 1974. *Biochemistry* 13:5017-5022
17. Dijkstra, B. W., Kalk, K. H., Hol, W. G. J., Drenth, J. 1981. *J. Mol. Biol.* 147:97-123
18. De Haas, G. H., Bonsen, P. P. M., Pieterson, W. A., van Deenen, L. L. M. 1971. *Biochim. Biophys. Acta* 239:252-266
19. Arnone, A., Bier, C. J., Cotton, F. A., Day, V. W., Hazen, E. E. Jr., Richardson, D. C., Richardson, J. S., Yonath, A. 1971. *J. Biol. Chem.* 246:2302-2316
20. Cotton, F. A., Hazen, E. E. Jr., Legg, M. J. 1979. *Proc. Natl. Acad. Sci. USA* 76:2551-2555
21. Hardman, K. D., Ainsworth, C. F. 1972. *Biochemistry* 11:4910-4919

22. Becker, J. W., Reeke, G. N. Jr., Wang, J. L., Cunningham, B. A., Edelman, G. M. 1975. *J. Biol. Chem.* 250:1513-1524

23. Kalb, A. J., Levitzki, A. 1968. *Biochem. J.* 109:669-672

24. Stuart, D. I., Acharya, K. R., Walker, N. P. C., Smith, S. G., Lewis, M., Phillips, D. C. 1986. *Nature* 324:84-87

25. Mitani, M., Harushima, Y., Kuwajima, K., Ikeguchi, M., Sugai, S. 1986. *J. Biol. Chem.* 261:8824-8829

26. Hiraoka, Y., Sugai, S. 1984. *Int. J. Peptide Protein Res.* 23:535-542

27. Musci, G., Berliner, L. J. 1985. *Biochemistry* 24:6945-6948

28. Vyas, N. K., Vyas, M. N., Quiocho, F. A. 1987. *Nature* 327:635-638

29. Kretsinger, R. H., Nockolds, C. E. 1973. *J. Biol. Chem.* 248:3313-3326

30. Moews, P. C., Kretsinger, R. H. 1975. *J. Mol. Biol.* 91:201-228

31. Swain, A. 1988. Ph.D. Thesis. University of South Carolina

32. Declercq, J.-P., Tinant, B., Parello, J., Etienne, G., Huber, R. 1988. *J. Mol. Biol.* 202:349-353.

33. Szebenyi, D. M. E., Obendorf, S. K., Moffat, K. 1981. *Nature* 294:327-332

34. Szebenyi, D. M. E., Moffat, K. 1986. *J. Biol. Chem.* 261:8761-8777

35. Szebenyi, D. M. E., Moffat, K. 1987. *Methods Enzymol.* 139:585-610

36. Babu, Y. S., Sack, J. S., Greenhough, T. J., Bugg, C. E., Means, A. R., Cook, W. J. 1985. *Nature* 315:37-40

37. Babu, Y. S., Bugg, C. E., Cook, W. J. 1987. *Methods Enzymol.* 139:632-642

38. Babu, Y. S., Bugg, C. E., Cook, W. J. 1988. *J. Mol. Biol.* In press

39. Herzberg, O., James, M. N. G. 1985. *Nature* 313:653-659

40. Herzberg, O., Moult, J., James, M. N. G. 1987. *Methods Enzymol.* 139:610-632

41. Herzberg, O., James, M. N. G. 1988. *J. Mol. Biol.* 202. In press

42. Sundaralingam, M., Bergstrom, R., Strasburg, G., Rao, S.T., Roychowdhury, P., Greaser, M., Wang, B.C. 1985. *Science* 227:945-948

43. Satyshur, K. A., Rao, S. T., Pyzalska, D., Drendel, W., Greaser, M., Sundaralingam, M. 1988. *J. Biol. Chem.* 263:1628-1647

44. Kretsinger, R. H. 1980. *CRC Crit. Rev. Biochem.* 8:119-174

45. Einspahr, H., Bugg, C. E. 1980. *Acta Cryst.* B36:264-271

46. Einspahr, H., Bugg, C. E. 1981. *Acta Cryst.* B37:1044-1052

47. Einsphar, H., Bugg, C. 1984. In *Metal Ions Biological Systems*, ed. H. Sigel, Vol. 17, pp. 51-97. New York and Basel: Marcel Dekker, Inc.

48. Herzberg, O., Moult, J., James, M. N. G. 1986. *J. Biol. Chem.* 261:2638-2644

49. Huxley, H. E. 1973. *Cold Spring Harb. Symp. Quant. Biol.* 37:361-376

50. Leavis, P. C., Gergely, J. 1984. *CRC Crit. Rev. Biochem.* 16:235-305

51. Zot, A. S., Potter, J. D. 1987.*Ann. Rev. Biophys. Biophys. Chem.* 16:535-559

52. Kay, C. M., McCubbin, W. D., Sykes, B. D. 1987. *Biopolymers* 26:S123-S144

53. Greaser, M. L., Gergely, J. 1973. *J. Biol. Chem.* 248:2125-2133

54. Potter, J. D., Gergely, J. 1975. 250:4628-4633

55. Grabarek, Z., W. Drabikowski. 1981. *J. Biol. Chem.* 256:13121-13127

56. Leavis, P. C., Rosenfeld, S. S., Gergely, J. 1978. *J. Biol. Chem.* 253:5452-5459

57. Talbot, J. A., Hodges, R. S. 1981. *J. Biol. Chem.* 256:2798-2802

58. Syska, H., Wilkinson, J. M., Grand, R. J. A., Perry, S. V. 1976. *Biochem. J.* 153:375-387

59. Weeks, R. A., Perry, S. V. 1978. *Biochem. J.* 173:449-457

60. van Eyk, J. E., Hodges, R. S. 1988. *J. Biol. Chem.* 263:1726-1732

61. Drabikowski, W., Dalgarno, D. C., Levine, B. A., Gergely, J., Grabarek, Z., Leavis, P. C. 1985. *Eur. J. Biochem.* 151:17-28

62. Cachia, P. J., Gariépy, J., Hodges, R. S. 1985. In *Calmodulin Antagonists and Cellular Physiology*, eds. H. Hidaka, D. J. Hartshorne, Chapter 5, pp. 63-88, New York: Academic Press Inc.

63. Levine, B. A., Coffman, D. M. D., Thornton, J. M. 1977. *J. Mol. Biol.* 115:743-760

64. Gariépy, J., Hodges, R. S. 1983. *FEBS Letters* 160:1-6

65. Ingraham, R. H., Swenson, C. A. 1984. *J. Biol. Chem.* 59:9544-9548

66. Wang, C.-K., Cheung, H. C. 1985. *Biophys. J.* 48:727-739

67. Haselgrove, J. C. 1973. *Cold Spring Harb. Symp. Quant. Biol.* 37:341-352

68. Wakabayashi, T., Huxley, H. E., Amos, L. A., Klug, A. 1975. *J. Mol. Biol.* 93:477-497

69. Chalovich, J. M., Chock, P. B., Eisenberg, E. 1981. *J. Biol. Chem.* 256:575-578

70. Potter, J. D., Seidel, J. C., Leavis, P., Lehrer, S. S., Gergely, J. 1976. *J. Biol. Chem.* 251:7551-7556

71. Ellis, P. D. Strang, P., Potter, J. D. 1984. *J. Biol. Chem.* 259:10348-10356

72. Means, A. R., Tash, J. S., Chafouleas, J. G. 1982. *Physiol. Rev.* 62:1-39

73. Klee, C. B., Crouch, T. H., Richman, P. G. 1980. *Ann. Rev. Biochem.* 49:489-515

74. Lin, Y. M., Liu, Y. P., Cheung, W. Y. 1974. *J. Biol. Chem.* 249:4943-4954

75. Teo, T. S., Wang, J. H. 1973. *J. Biol. Chem.* 248:5950-5955

76. Hathaway, D. R., Adelstein, R. S. 1979. *Proc. Natl. Acad. Sci. USA* 76:1653-1657

77. Dabrowska, R., Sherry, J. M. F., Aromatorio, D. K., Hartshorne, D. J. 1978. *Biochemistry* 17:253-258

78. Stewart, A. A., Ingebritsen, T. S., Manalan, A., Klee, C. B., Cohen, P. 1982. *FEBS Letters* 137:80-84

79. Lynch, T. J., Cheung, W. Y. 1979. *Arch. Biochem. Biophys.* 194:165-170

80. Hanahan, D. J., Taverna, R. D., Flynn, D. D., Ekholm, J. E. 1978. *Biochem. Biophys. Res. Comm.* 84:1009-1015

81. Brostrom, M. A., Brostrom, C. O., Wolff, D. J. 1978. *Arch. Biochem. Biophys.* 191:341-350

82. Westcott, K. R., La Porte, D. C., Storm, D. R. 1979. *Proc. Natl. Acad. Sci. USA* 76:204-208

83. Cohen, P., Burchell, A., Foulkes, J. G., Cohen, P. T. W., Vanaman, T.C., Nairn, A.C. 1978. *FEBS Letters* 92:287-293

84. Anderson, J. M., Cormier, M. J. 1978. *Biochem. Biophys. Res. Comm.* 84:595-602

85. Klee, C. B., Vanaman, T. C. 1982. *Adv. Prot. Chem.* 35:213-321

86. La Porte, D.C., Wierman, B. M., Storm, D. R. 1980. *Biochemistry* 19:3814-3819

87. Tanaka, T., Hidaka, H. 1980. *J. Biol. Chem.* 255:11078-11080

88. Tanaka, T., Hidaka, H. 1981. *Biochem. Biophys. Res. Comm.* 101:447-453

89. Klevit, R. E., Dalgarno, D. C., Levine, B. A., Williams, R. J. P. 1984. *Eur. J. Biochem.* 139:109-114

90. Krebs, J., Buerkler, J., Guerini, D., Brunner, J., Carafoli, E. 1984. *Biochemistry* 23:400-403

91. Follenius, A., Gerard, D. 1984. *Biochem. Biophys. Res. Comm.* 119:1154-1160

92. Dalgarno, D. C., Klevit, R. E., Levine, B. A., Scott, G. M. M., Williams, R. J. P., Gergely, J., Grabarek, Z., Leavis, P. C., Grand, R. J. A., Drabikowski, W. 1984. *Biochim. Biophys. Acta* 791:164-172

93. Burger, D., Cox, J. A., Comte, M., Stein, E. A. 1984. *Biochemistry* 23:1966-1971

94. Jarrett, H. W. 1984. *J. Biol. Chem.* 259:10136-10144

95. Zimmer, M., Hofmann, F. 1987. *Eur. J. Biochem.* 164:411-420

96. Seamon, K. B. 1980. *Biochemistry* 19:207-215

97. Crouch, T. H., Klee, C. B. 1980. *Biochemistry* 19:3692-3698

98. Wolff, D. J., Poirier, P. G., Brostrom, C. O., Brostrom, M. A. 1977. *J. Biol. Chem.* 252:4108-4117

99. Martin, S. R., Andersson Teleman, A., Bayley, P. M., Drakenberg, T., Forsen, S. 1985. *Eur. J. Biochem.* 151:543-550

100. Teleman, A., Drakenberg, T., Forsén, S. 1986. *Biochim. Biophys. Acta* 873:204-213

101. Forsén, S., Vogel, H. J., Drakenberg, T. 1986. In *Calcium and Cell Function*, Vol. VI, pp. 113-157, New York: Academic Press Inc.

102. Means, A. R. 1988. *Recent Progresses in Hormone Research* 44:223-286

103. Baron, G, Demaille, J., Dutruge, E. 1975. *FEBS Letters* 56:156-160

104. Blum, H. E., Lehky, P., Kohler, L., Stein, E. A., Fischer, E.H. 1977. *J. Biol. Chem.* 252:2834-2838

105. Pechère, J. -F., Derancourt, J., Haiech, J. 1977. *FEBS Letters* 75:111-114

106. Haiech, J., Derancourt, J., Pechère, J. -F., Demaille, J.G. 1979. *Biochemistry* 18:2752-2758

107. Gillis, J. M., Thomason, P., Lefevre, I., Kretsinger, R. H. 1982. *J. Muscle Res. Cell. Motil.* 3:377-398

108. Wnuk, W., Cox, J. A., Stein, E. A. 1982. In *Calcium and Cell Function*, Vol. II, ed. W. Y. Cheung, pp. 243-278, New York: Academic Press Inc.

109. Perchère, J. F. 1977. In *Calcium binding proteins and Calcium Function*. eds. R. H. Wasserman et al, pp. 212-221, New York: North-Holland

110. Permyakov, E. A., Yarmolenko, V. V., Emelyanenko, V. I., Burstein, E. A., Closset, J., Gerday, C. 1980. *Eur. J. Biochem.* 109:307-315

111. Grandjean, J., Laszlo, P., Gerday, C. 1977. *FEBS Letters* 81:376-380

112. Taylor, A. N., 1983. In *Calcium Binding Proteins* , eds. B. de Bernard et al, pp. 207-213, Amsterdam: Elsevier Science Publishers

113. Wasserman, R. H., Fullmer, C.S. 1982. In *Calcium and Cell Function* , Vol. II, ed. W.Y. Cheung, pp. 175-216, New York: Academic Press Inc.

114. Wasserman, R. H., Shimura, F. Meyer, S. A., Fullmer, C. S. 1983. In *Calcium Binding* Proteins, ed. B. de Bernard et al, pp. 183-205, Amsterdam: Elsevier Science Publishers

115. Levine, B A., Williams, R. J. P. 1982. In *Calcium and Cell Function*, Vol. II, ed. W. Y. Cheung, pp. 1-38, New York: Academic Press Inc.

116. Dalgarno, D C., Levine, B. A., Williams, R. J. P., Fullmer, C. S., Wasserman, R. H. 1983. *Eur. J. Biochem.* 137:523-529

117. Chiba, K., Mohri, T. 1987. *Biochemistry* 26:711-715

118. Shelling, J. G., Sykes, B. D., O'Neil, J. D. J., Hofmann, T. 1983. *Biochemistry* 22:2649-2654

119. O'Neil, J. D. J., Dorrington, K. J., Hofmann, T. 1984. *Can. J. Biochem.Cell Biol.* 62:434-442

120. Bryant, D. T. W., Andrews, P. 1984. *Biochem. J.* 219:287-292

121. Shelling, J. G., Sykes, B. D. 1985. *J. Biol. Chem.* 260:8342-8347

122. Wilkinson, J. M. 1976. *FEBS Letters* 70:254-256

123. Reinach, F. C., Karlsson, R. 1988. *J. Biol. Chem.* 263:2371-2376

124. Watterson, D. M., Sharief, F., Vanaman, T. C. 1980. *J. Biol. Chem.* 255:962-975

125. Watterson, D., Burgess, W. H., Lukas, T., Iverson, D., Marshak, D. R. et al. 1984. *Adv. Cyclic Nucleotide Protein Phosphorylation Research* 16:205-226

126. Coffee, C. J., Bradshaw, R. A. 1973. *J. Biol. Chem.* 248:3305-3312

127. Fullmer, C S., Wasserman, R. H. 1981. *J. Biol. Chem.* 256:5669-5674

128. Tufty, R. M., Kretsinger, R. H. 1975. *Science* 187:167-169

129. Kobayashi, T., Takagi, T., Konishi, K., Ohnishi, K., Watanabe, Y. 1988. *Eur. J. Biochem.* 174:579-584

130. Hardin, S. H., Keast, M. J., Hardin, P. E., Klein, W. H. 1987. *Biochemistry* 26:3518-3523

131. Putkey, J. A., Ts'ui, K. F., Tanaka, T., Lagacé, L, Stein, J. P., Lai, E. C., Means, A. R. 1983. *J. Biol. Chem.* 258:11864-11870

132. Herzberg, O., Hayakawa, K., James, M. N. G. 1984. *J. Mol. Biol.* 172:345-346

133. Cook, W. J., Sack, J. S. 1983. *Methods Enzymol.* 102:143-147

134. Hendrickson, W. A., Konnert, J H. 1980. In *Computing in Crystallography*, eds. R. Diamond, S. Ramaseshan, K. Venkatesan, pp. 13.01-13.23, Indian Academy of Sciences, Inter. Union of Crystallography, Bangalore, India

135. Chambers, J. L., Stroud, R. M. 1979. *Acta Cryst.* B35:1861-1874

136. Blevins, R. A., Tulinsky, A. 1985. *J. Biol. Chem.* 260:4264-4275

137. Tsukada, H., Blow, D M. 1985. *J. Mol. Biol.* 184:703-711

138. Fujinaga, M., Sielecki, A. R., Read, R. J., Ardelt, W., Laskowski, M. Jr., James, M. N. G. 1987. *J. Mol. Biol.* 195:397-418

139. Read, R. J. 1986. *Acta Cryst*. A42:140-149

140. van Eerd, J. -P, Takahashi, K. 1976. *Biochemistry* 15:1171-1180

141. Wnuk, W., Schoechlin, M., Stein, E. A. 1984. *J. Biol. Chem.* 259:9017-9023

142. Strynadka, N. C. J., James, M. N.G. 1988. *Proteins: Structure, Function and Genetics* 3:1-17

143. Putkey, J. A., Ono, T., VanBerkum, M. F. A., Means, A. R. 1988. *J. Biol. Chem.* 263:11242-11249

144. Schutt, C. 1985. *Nature* 315:15-15

145. Heidorn, D. B., Trewhella, J. 1988. *Biochemistry* 27:909-915

146. Hubbard, S. R., Hodgson, K. O., Doniach, S. 1988. *J. Biol. Chem.* 263:4151-4158

147. Seaton, B. A., Head, J. F., Engelman, D.M., Richards, F. M. 1985. *Biochemistry* 24:6740-6743

148. Fujisawa, T., Ueki, T., Inoko, Y. 1987. *J. Appl. Cryst.* 20:349-355

149. Kretsinger, R. H. 1972. *Nature (London)* 240:85-88

150. Suguna, K., Bott, R. R., Padlan, E. A., Subramanian, E., Sheriff, S., Cohen, G. H., Davies, D. R. 1987. *J. Mol. Biol.* 196:877-900

151. Saenger, W. 1984. In *Principles of Nucleic Acid Structure*, ed. C. R. Cantor, pp. 55-69, New York: Springer-Verlag

152. Shannon, R. D., Prewitt, C. T. 1988. *Acta Cryst.* B25:925-946

153. Rodisiler, P. F., Amma, E. L. 1982. *J. Chem. Soc. Chem. Commun.* 182-184

154. Weeds, A. G., McLachlan, A. D. 1974. *Nature (London)* 252:646-649

155. Nemethy, G., Printz, M. P. 1972. *Macromolecules* 5:755-758

156. Baker, E. N., Hubbard, R. E. 1984. *Prog. Biophys. Molec. Biol.* 44:97-179

157. Herzberg, O., James, M. N. G. 1985. *Biochemistry* 24:5298-5302

158. Dalgarno, D. C., Klevit, R. E., Levine, B. A., Williams, R. J. P., Dobrowolski, Z., Drabikowski, W. 1984. *Eur. J. Biochem.* 138:281-289

159. Ikura, M., Minowa, O., Hikichi, K. 1985. *Biochemistry* 24:4264-4269

160. Teleman, O., Drakenberg, T., Forsén, S., Thulin, E. 1983. *Eur. J. Biochem.* 134:453-457

161. Linse, S., Brodin, P., Drakenberg, T., Thulin, E., Sellers, P., Elmdén, K., Grundström, T., Forsén, S. 1987. *Biochemistry* 26:6723-6735

162. Herzberg, O., Moult, J., James, M. N. G. 1986. In *Calcium and the Cell, Ciba Foundation Symposium 122*, eds. D. Evered, J. Whelan, pp. 120-144, Chichester, U.K.: John Wiley & Sons Ltd.

163. Vogel, H. J., Drakenberg, T. D., Forsén, S., O'Neil, J. D. J., Hoffmann, T. 1985. *Biochemistry* 24:3870-3876

164. Fullmer, C. S., Wasserman, R. H. 1980. In *Calcium Binding Proteins: Structure and Function*, eds. F. L. Siegel, E. Carafoli, R. H. Kretsinger, D. H. MacLennan, R. H. Wasserman, pp. 363-370, New York: Elsevier/North-Holland

165. Wang, C.-L. A., Gergely, J. 1986. *Eur. J. Biochem.* 154:225-228

166. Grabarek, Z., Leavis, P. C., Gergely, J. 1986. *J. Biol. Chem.* 261:608-613

167. Newton, D. L., Oldewurtel, M. H., Krinks, M. H., Shiloach, J., Klee, C. B. 1984. *J. Biol. Chem.* 259:4419-4426

168. Drabikowski, W., Grabarek, Z., Barylko, B. 1977. *Biochim. Biophys. Acta* 490:216-224

169. Persechini, A., Kretsinger, R. H. 1988. *J. Biol. Chem.* 263:12175-12178

170. Persechini, A., Kretsinger, R. H. 1988. *J. Cardiovasc. Pharmacol.* In press

171. Weber, P. C., Salemme, F. R. 1980. *Nature* 287:82-85

172. Sheridan, R. P., Levy, R. M., Salemme, F. R. 1982. *Proc. Natl. Acad. Sci. USA* 79:4545-4549

173. Sekharudu, Y. C., Sundaralingam, M. 1988. *Protein Engineering* 2:139-146

174. Aulabaugh, A., Niemczura, W. P., Blundell, T. L., Gibbons, W. A. 1984. *Eur. J. Biochem.* 143:409-418

175. Krebs, J., Carafoli, E. 1982. *Eur. J. Biochem.* 124:619-627

176. Guerini, D., Krebs, J., Carafoli, E. 1987. *Eur. J. Biochem.* 170:35-42

177. Klevit, R. E., Blumenthal, D. K., Wemmer, D. E., Krebs, E. G. 1985. *Biochemistry* 24:8152-8157

178. Walsh, M., Stevens, F. C. 1978. *Biochemistry* 17:3924-3930

179. Thiry, P., Vandermeers, A., Vedermeers-Piret, M.-C., Rathe, J., Christophe, J. 1980. *Eur. J. Biochem.* 103:409-414

180. Tanaka, T., Ohmura, T., Hidaka, H. 1983. *Pharmacology* 26:249-257

181. Blumenthal, D. K., Takio, K., Edelman, A. M., Charbonneau, H., Titani, K., Walsh, K. A., Krebs, E. G. 1985. *Proc. Natl. Acad. Sci. USA* 82:3187-3191

182. Buschmeier, B., Meyer, H. E., Mayr, G. W. 1987. *J. Biol. Chem.* 262:9454-9462

183. Prozialeck, W. C., Weiss, B. 1982. *J. Pharmacol. Exper. Ther.* 222:509-514

184. Erickson-Viitanen, S., O'Neil, K. T., DeGrado, W. F. 1987. In *Protein Engineering*, eds. D. L. Oxender, C. F. Fox, pp. 201-211, New York: Alan R. Liss, Inc.

185. Gariépy, J., Hodges, R. S. 1983. *Biochemistry* 22:1586-1594

186. O'Neil, K. T., DeGrado, W. F. 1985. *Proc. Natl. Acad. Sci. USA* 82:4954-4958

187. Gresh, N. 1987. *Molec. Pharmacol.* 31:617-622

188. Craig, T. A., Watterson, D. M., Prendergast, F. G., Haiech, J., Roberts, D. M. 1987. *J. Biol. Chem.* 262:3278-3284

189. Jackson, A. E., Carraway, K. L. III, Puett, D., Brew, K. 1986. *J. Biol. Chem.* 221:12226-12232

190. Manalan, A. S., Klee, C. B. 1987. *Biochemistry* 26:1382-1390

191. Winkler, M. A., Fried, V. A., Merat, D. L., Cheung, W. Y. 1987. *J. Biol. Chem.* 262:15466-15471

192. Faust, F. M., Slisz, M., Jarrett, H. W. 1987. *J. Biol. Chem.* 262:1938-1941

193. Mann, D. M., Vanaman, T. C. 1988. *J. Biol. Chem.* 263:11284-11290

194. Warshel, A. 1978. *J. Phys. Chem.* 83:1640-1652

195. Gilson, M. K., Honig, B. 1988. *Proteins: Structure, Function, and Genetics* 4:7-18

196. Sundaralingam, M., Drendel, W., Greaser, M. 1985. *Proc. Natl. Acad. Sci. USA* 82:7944-7947

197. Sundaralingam, M., Sekharudu, Y. C., Yathindra, N., Ravichandran, V. 1987. *Proteins: Structure, Function, and Genetics* 2:64-71

198. Drakenberg, T., Forsén, S., Thulin, E., Vogel, H. J. 1987. *J. Biol. Chem.* 262:672-678

Table 3.1

Calcium Binding Proteins Crystal Structures

Protein	Resolution (Å)	Ca^{2+}-Binding Characteristics: No. Ca^{2+} Sites	Coordinating Peptide[a]	Coordination Number	$K_d^{Ca^{2+}}$ (M)	Role of Ca^{2+} ion	References
bovine β-trypsin	1.9	1	C	6	$4x10^{-4}$	stabilizing	2,3,4
S. griseus trypsin	1.7	1	D	7	-	stabilizing	5,6,7
subtilisin Novo	2.1	1	D	7	-	stabilizing	8
subtilisin Carlsberg	1.2,1.8	1	D	7	-	stabilizing	9,10,11,12
thermolysin	1.6	4	D	6,7	$2x10^{-5}$ $1x10^{-6}$	stabilizing	13,14,15,16
phospholipase A_2	1.7	1	D	7	$2.5x10^{-4}$	stabilizing (catalytic)	17,18
staphylococcal nuclease	1.5	1	D	6	$\sim 10^{-3}$	(catalytic)	19,20
concanavalin A	1.75	1	D	7	$3x10^{-4}$	cell binding	21,22,23
α-lactalbumin	1.7	1	C	7	10^{-6}—10^{-7}	stabilizing cofactor	24,25,26,27
D-galactose binding prot.	1.9	1	D	7	-	-	28
carp Parv	1.6	2	C	7	$\sim 10^{-9}$	intracellular calcium buffer	29,30,31
pike Parv	1.9	2	C	7	10^{-8}	intracellular calcium buffer	32
bovine ICaBP	2.3	2	C	7	10^{-8}	calcium transport	33,34,35
bovine CaM	2.2	4	C	7	10^{-5}, 10^{-6}	intracellular enzyme regulation	36,37,38
avian TnC	2.0	4	C	7	10^{-5}, 10^{-7}	muscle contraction and regulation	39,40,41,42,43

[a] C denotes a continuous polypeptide that contains the Ca^{2+} ligands.
D indicates that the Ca^{2+} ligands come from several separated segments of polypeptide.

Table 3.2

Primary Sequences

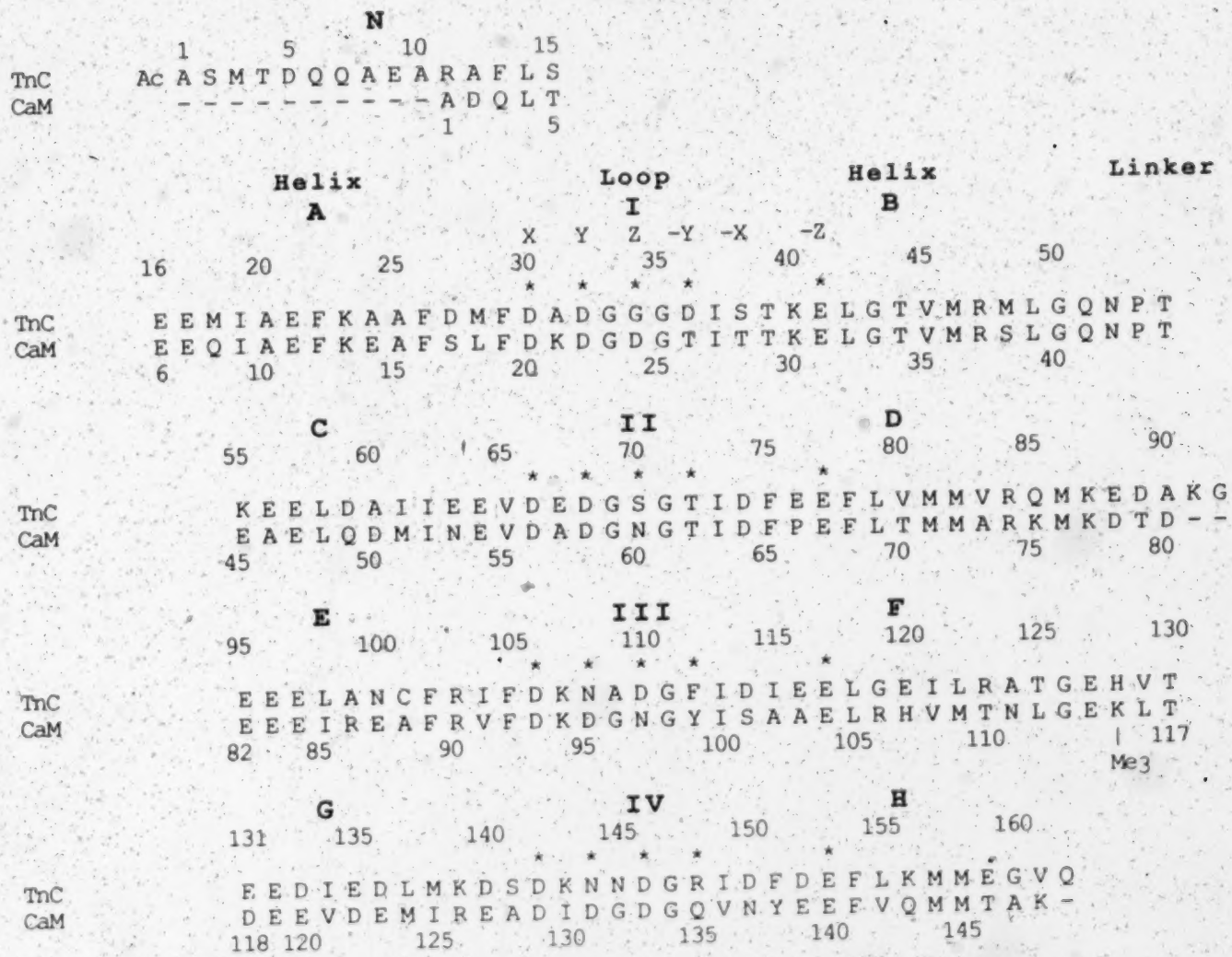

N
1 5 10 15
TnC Ac A S M T D Q Q A E A R A F L S
CaM - - - - - - - - - - A D Q L T
1 5
Helix A
Loop I
Helix B
Linker
X Y Z -Y -X -Z
16 20 25 30 35 40 45 50
TnC E E M I A E F K A A F D M F D A D G G G D I S T K E L G T V M R M L G Q N P T
CaM E E Q I A E F K E A F S L F D K D G D G T I T T K E L G T V M R S L G Q N P T
6 10 15 20 25 30 35 40
C II D
55 60 65 70 75 80 85 90
TnC K E E L D A I I E E V D E D G S G T I D F E E F L V M M V R Q M K E D A K G K S
CaM E A E L Q D M I N E V D A D G N G T I D F P E F L T M M A R K M K D T D - - - S
45 50 55 60 65 70 75 80
E III F
95 100 105 110 115 120 125 130
TnC E E E L A N C F R I F D K N A D G F I D I E E L G E I L R A T G E H V T
CaM E E E I R E A F R V F D K D G N G Y I S A A E L R H V M T N L G E K L T
82 85 90 95 100 105 110 117
Me3
G IV H
131 135 140 145 150 155 160
TnC E E D I E D L M K D S D K N N D G R I D F D E F L K M M E G V Q
CaM D E E V D E M I R E A D I D G D G Q V N Y E E F V Q M M T A K -
118 120 125 130 135 140 145

Table 2 contd.

```
                        A                 B
          1         5         10        15        20        25        30        35
Parv[a]   Ac A F A G V L N D A D I A A A L E A C K A A D S F D H K A F F A K V G L T S K S

                        C                 I                 D
                  40        45        50        55        60        65        70        75
                                      *   *   *   *   *   *
Parv              A D D V K K A F A I I D Q D K S G F I E E D E L K L F L Q N F K A D A R A L T

                        E                 II                F
                  79 80     85        90        95        100       105
                                      *   *   *   *       *
Parv              D G E T K T F L K A G D S D G D G K I G V D E F T A L V K A

                        A                 I                 B
          1         5         10        15        20        25        30        35        40        45
                                     *A   *   *P   *       *
ICaBP     K S P E E L K G I F E K Y A K E G D N Q L S K E E L K L L L Q T E F P S L L K G P S T

                        C                 II                D
                            50        55        60        65        70        75
                                      *   *   *   *       *
ICaBP             L D E L F E E L D K N G D G E V S F E E F Q V L V K K I S Q
```

[a] Helix A extends from Asp8 to Ala17; helix B from His26 to Val33 (31).

* The residues that form the calcium coordination sphere, labelled X, Y, Z, -Y, -X, -Z, are denoted by the asterisk above those residues in each loop.

The amino-acid sequences are given in the one letter code. The sources of the sequences are: chicken skeletal TnC (122,123); bovine brain CaM (124,125); carp Parv (126); bovine ICaBP (127).

Table 3.3

Summary of Crystallographic Data

	Turkey sTnC	Chicken sTnC	CaM	Parv	ICaBP
Crystallization conditions:					
	40% $(NH_4)_2SO_4$	43% $(NH_4)_2SO_4$	45-60% MPD	70-85% $(NH_4)_2SO_4$	70-80% $(NH_4)_2SO_4$
	100 mM NaOAc	50 mM NaOAc	50 mM cacodylate	2 mM phosphate	20 mM Tris
	2.5-5% PEG200	3% PEG200	4 mM $CaCl_2$		
	10 mM $CaCl_2$	5 mM $MgCl_2$			
	pH 4.9-5.0	pH 5.1	pH 5.2-5.8	pH 6.8	pH 8.9
	Ref. 132	Ref. 43	Ref. 133	Ref. 29,30	Ref. 33
Crystallographic Refinement Information:					
	Ref.41	Ref.43	Ref. 38	Ref. 31	Ref. 34
Final R[a]	.155	.172	.175	.187	.178
Resolution (Å)	10.0-2.0	8.0-2.0	5.0-2.2	10.0-1.6	9.0-2.3
Cutoff	$I \geq 2\sigma(I)$	$I \geq 2\sigma(I)$	$I \geq 2.5\sigma(I)$	$I \geq 3\sigma(I)$	not given
No. reflections	8054	8100	6685	10879	2350
No. protein atoms	1259	1248	1121	808	600
Ca^{2+} ions	2	2	4	2	2
No. solvent atoms	158	68	69	75	37 ($1SO_4^=$)
Residues not seen	1,2,162,67	1,2	1-4,148	-	
Deviations from ideal geometry[b]					
Bond distances	.019 (.017)	.04 (.03)	.016 (.020)	.020 (.02)	.016 (.02)

Table 3.3 contd.

Angle distances	.045 (.028)	.07 (.05)	.034 (.030)	.037 (.04)	.040 (.03)
Planar 1-4 distances	.049 (.034)	.11 (.10)	.044 (.040)	.060 (.05)	.061 (.06)
Transpeptide ω, (°) (deviations from 180°)	2.5 (3.5)	4.9 (5.0)	2.3 (5.0)	3.9 (3.0)	2.1 (3.0)
Estimated coord. error (Å)	~.30	~.17	~.15	~.15	~.25

[a] The agreement factor R is defined as $\Sigma||F_o|-|F_c||/\Sigma|F_o|$, where $|F_o|$ and $|F_c|$ are the observed and calculated structure factor amplitudes, respectively.

[b] The values given correspond to the root mean square deviations from ideal stereochemistry. The numbers in parentheses are the final input parameters to the refinement program that determine the relative weights of the corresponding restraints (134).

Table 3.4

Ca^{2+}-Binding Domain Comparisons[a]

	TnC-C	CaM-N	CaM-C	Parv
TnC-C	-	0.951[b]	0.817	0.827
CaM-N	197	-	0.751	0.865
CaM-C	238	238	-	0.855
Parv	228	203	223	-

[a] The Ca^{2+}-binding domains are defined in Table 3.2. The comparisons of the corresponding pairs of HLH units were carried out using a least-squares fitting computer program originally written by W. Bennett. Only main-chain atoms (N,C^{α},C,O) were included in the calculations; the atoms of the linking peptides were omitted. For TnC and CaM, the N- and C-terminal domains are denoted by an N or C following the name.

[b] The numbers in the upper triangle of this matrix are the rms deviations (in Å units) for each pairwise comparison. Those in the lower triangle are the number of pairs of atoms used in the least squares superposition. Atom pairs with a final deviation larger than 1.9 Å were considered non-equivalent and omitted from the corresponding comparison.

Table 3.5

Superpositions of Pairs of Loops

Protein	TnC-N	CaM-N	CaM-C	Parv	ICaBP[a]
TnC-C	3.068[b]	0.512	0.491	0.428	0.765
CaM-N		-	0.510	0.495	0.878
CaM-C			-	0.594	0.994
Parv				-	0.864

[a] For loop I of ICABP the least-squares superposition used only residues Gln22 to Glu27 plus all the residues in loop II.

[b] The numbers listed are the rms deviations in Ångstrom units. They result from the least-squares fitting of the 96 main chain atoms (N, C^{α}, C, O) of the loop regions of pairs of EF-hands (see footnote to Table 3.2).

Table 3.6

Interhelix Angles and Contact Distances

Helices	TnC		CaM		Parv[a]		ICaBP[a]	
	Angle (°)	Dist. (Å)	Angle (°)	Dist. (Å)	Angle (°)	Dist. (Å)	Angle (°)	Dist. (Å)
N/D	61	8.3	-	-				
A/B	134	10.0	96	*				
B/C	122	9.7	112	9.4				
C/D	146	11.8	84	*				
A/D	115	11.2	116	10.4				
B/D	53	11.3	-	-				
E/F	109	*	105	*	97	*	129	8.4
F/G	124	10.2	111	9.3	121	9.1	116	10.4
G/H	108	*	91	*	106	*	119	8.7
E/H	123	10.8	120	10.4	110	9.5	123	10.9

[a] For Parv and ICaBP, the interhelix angle and distance data have been aligned with the values for the C-terminal domains of TnC and CaM only for convenience of presentation. The standardized nomenclature for these helices is for Parv CDEF and for ICaBP ABCD.

* These interhelix distances have been omitted to indicate that the helix pairs concerned make contact only at the points where they enter or exit the Ca^{2+}-binding loop. Therefore such distances are not representative of the overall interhelical interactions and would be misleading (see Figure 3.1).

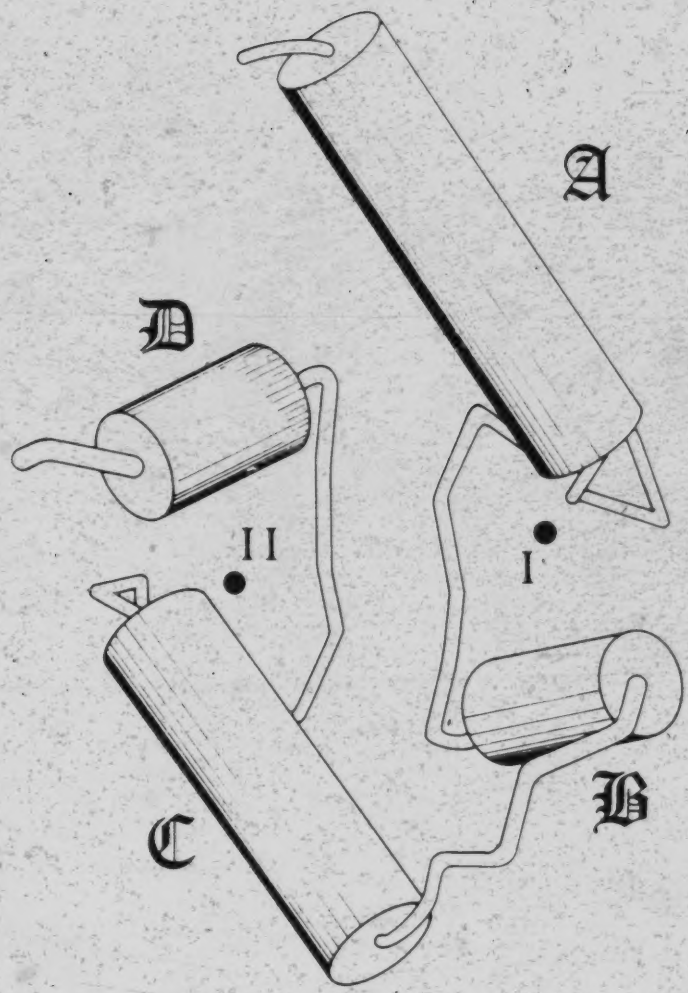
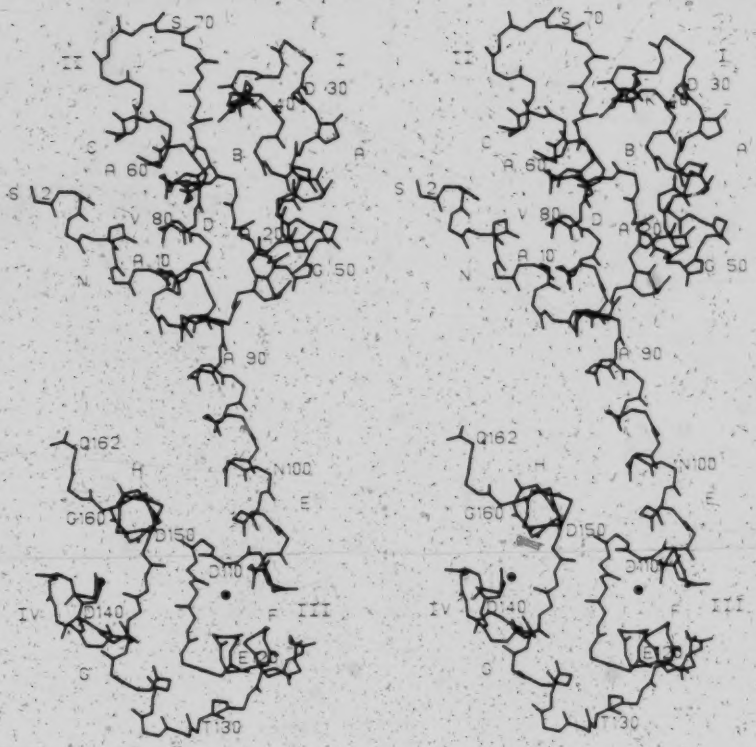
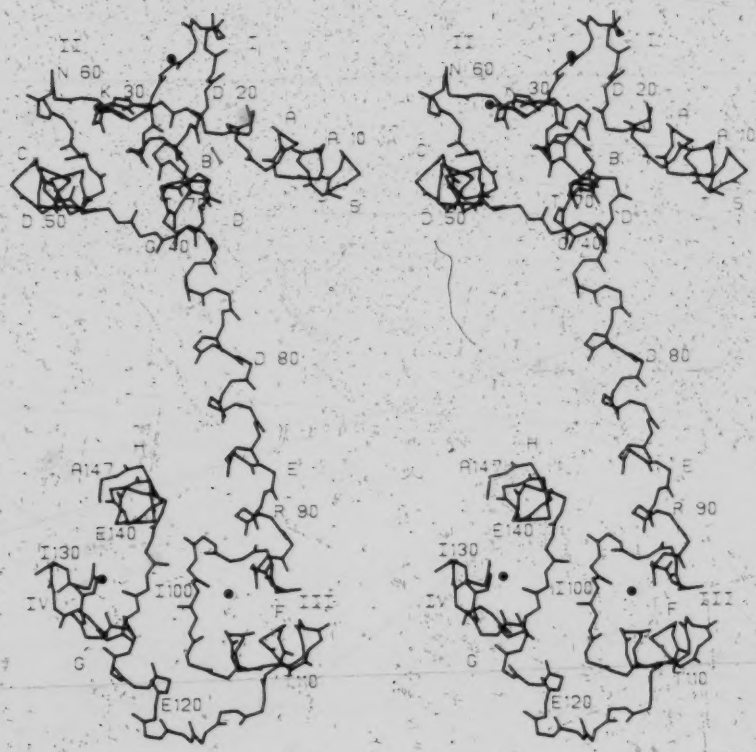


Fig. 3.1. A diagrammatic representation of a typical HLH Ca^{2+}-binding domain. There are two HLH structural units (helix A-loop I-helix B and helix C-loop II-helix D) that are related by a pseudo-two-fold rotation axis between the loops. Ca^{2+} ions are represented by filled circles (•).

a



b



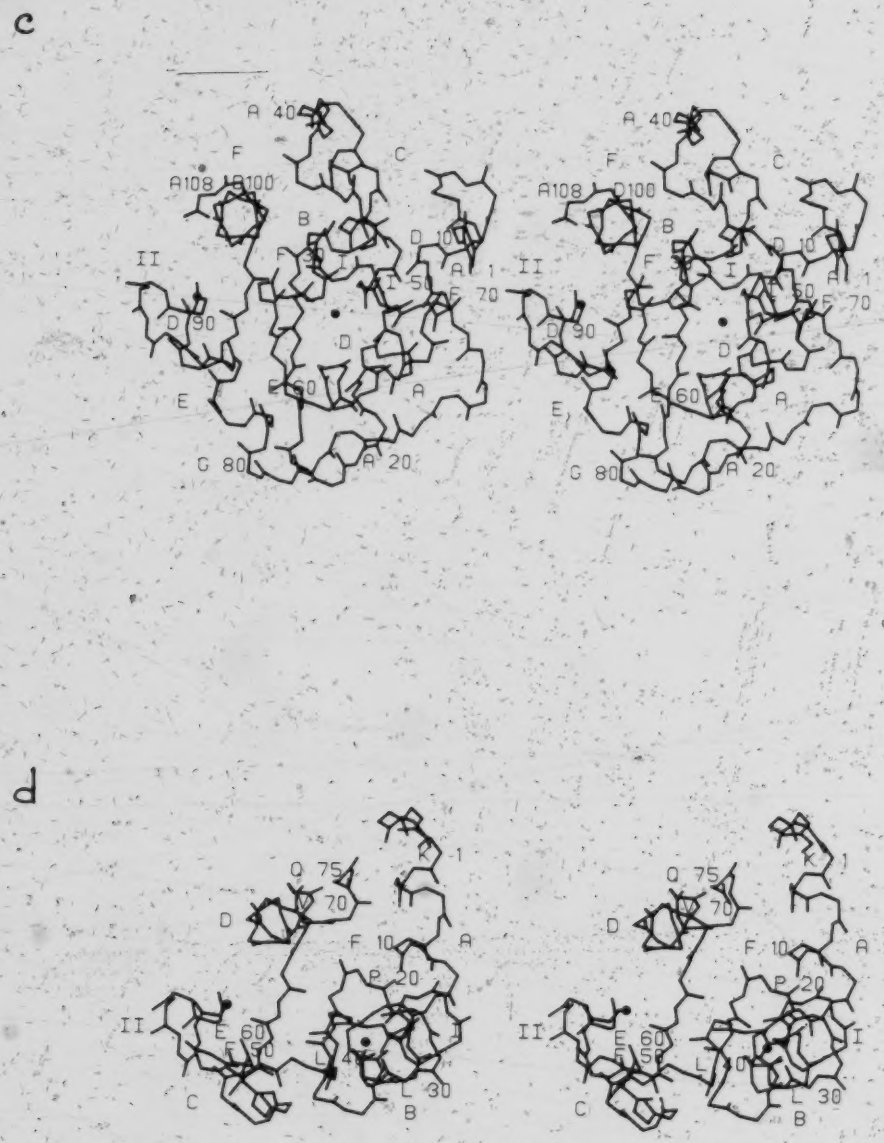


Fig. 3.2. Stereoscopic views of the four HLH proteins discussed in this review. Only the polypeptide chain atoms, N, C^{α}, C, O and the Ca^{2+} ions of each protein are represented. The Ca^{2+}-binding domains of CaM (C-terminal domain), Parv and ICaBP were structurally aligned to the C-terminal Ca^{2+}-binding domain of TnC and all four displayed from the same vantage point. (a) turkey skeletal muscle TnC. (b) bovine brain CaM. (c) carp Parv. (d) bovine ICaBP.

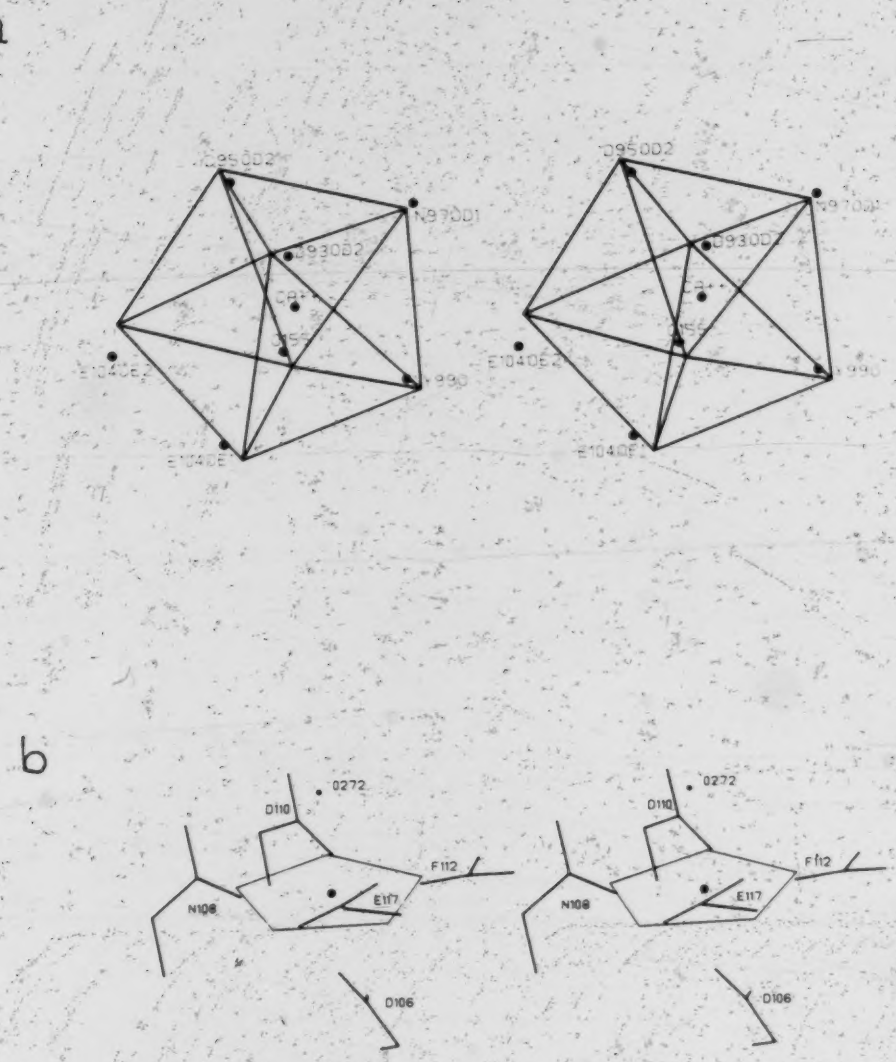
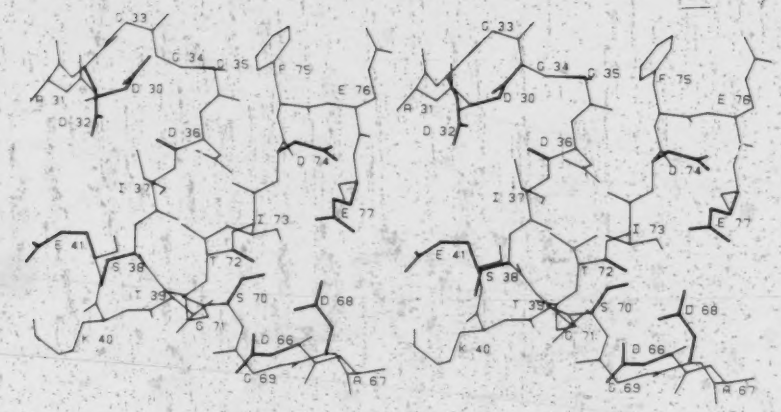
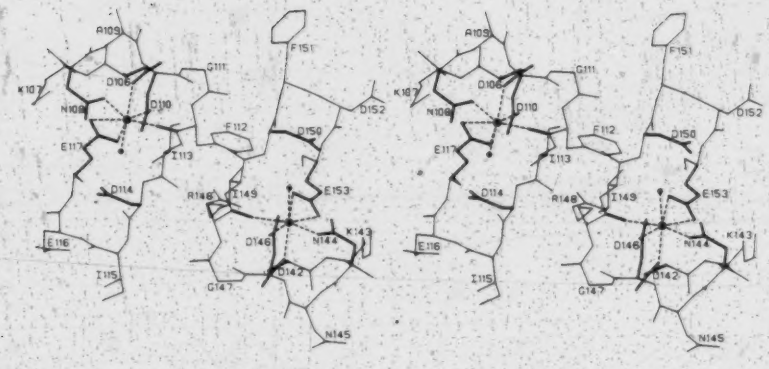


Fig. 3.3. (a) A stereoscopic representation of the oxygen atom ligands of Ca^{2+}-binding site III in CaM superimposed onto an ideal pentagonal bipyramid by a least-squares procedure. (b) The residues of the same Ca^{2+}-binding site showing the deviation of Glu104 $O^{\varepsilon 1}$ from coplanarity with the pentagonal equatorial plane. The other four oxygen ligands and the Ca^{2+} ion lie in the plane of the superposed pentagon.

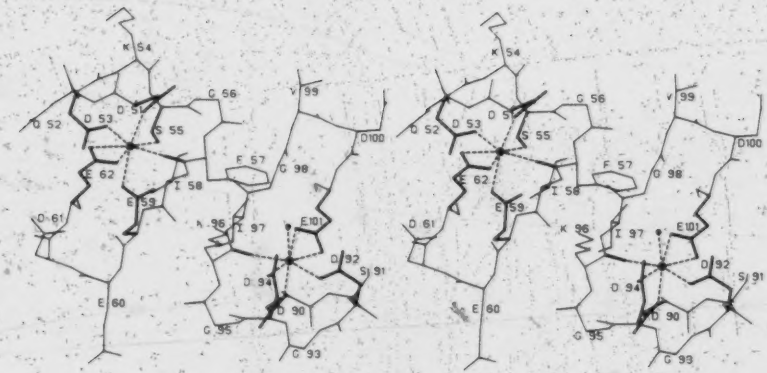
a


G 33
G 34
G 35
F 75
E 76
A 31
D 30
D 32
D 36
D 74
I 37
I 73
E 77
E 41
S 38
T 72
T 39
S 70
D 68
G 71
K 40
D 66
G 69
A 67

b


A109
F151
K107
D106
G111
D110
N108
D152
E112
D150
E117
I113
D114
R148
I149
E153
D146
N144
K143
E116
D142
I115
G147
N145

c


K 54
G 56
V 99
D 52
D 53
D 51
S 55
D100
E 57
G 98
E 62
I 58
E 59
D 61
K 96
I 97
E101
D 92
S 91
D 94
D 90
E 60
G 95
G 93

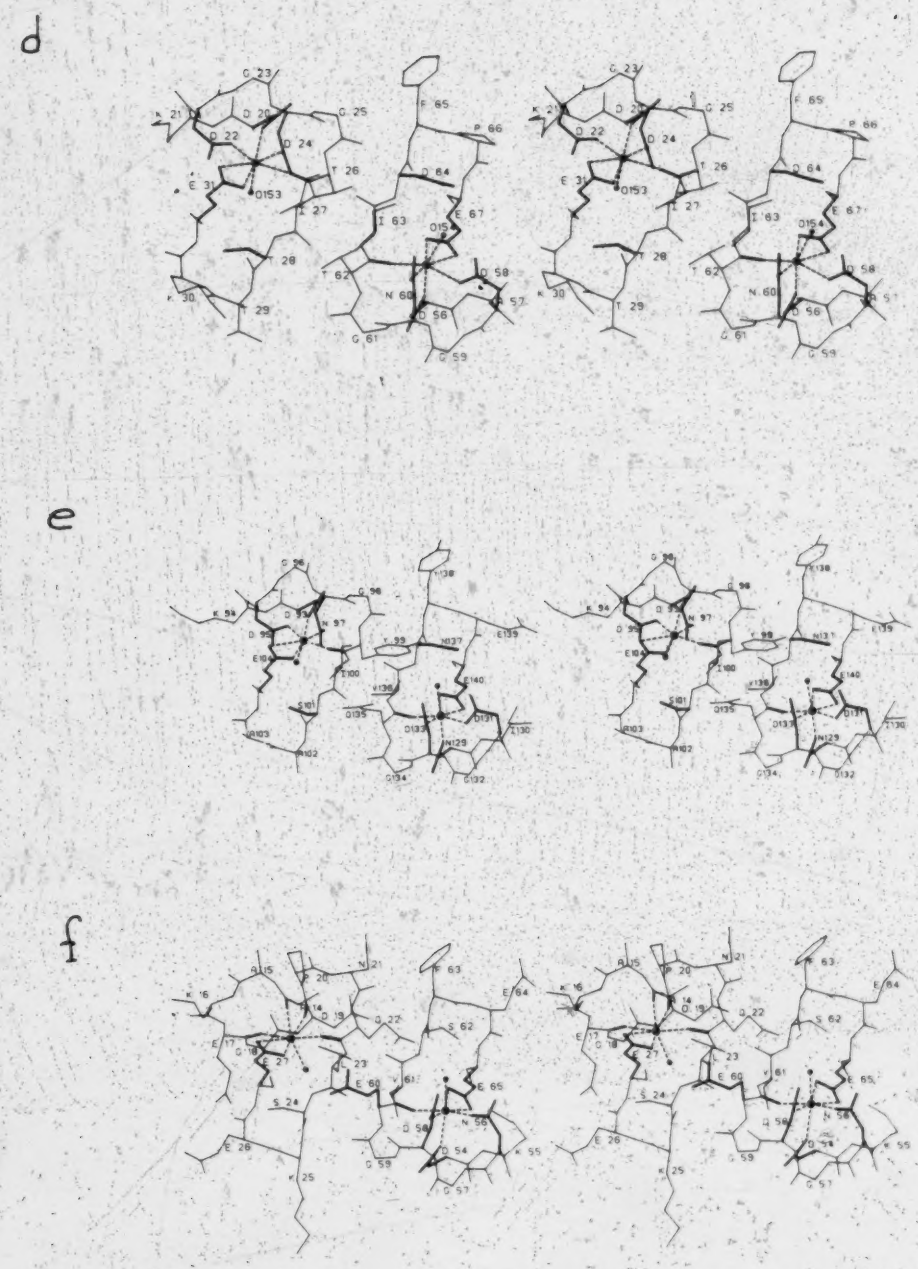


Fig. 3.4. Stereoscopic views of the pairs of 12 residue loops in each of the Ca^{2+}-binding loops of the four proteins. The Ca^{2+} ions are connected to their coordinating ligands by dashed lines. Residues in positions 1, 3, 5, 7 (C=O), 9 and 12 are represented by thick lines. (a) The Ca^{2+}-free loops I and II of TnC. (b) Loops III and IV of TnC. (c) Loops I and II of Parv. (d) Loops I and II of CaM. (e) Loops III and IV of CaM. (f) Loops I and II of ICaBP. There are two insertions in the first part of loop I in ICaBP that contribute to its altered conformation.

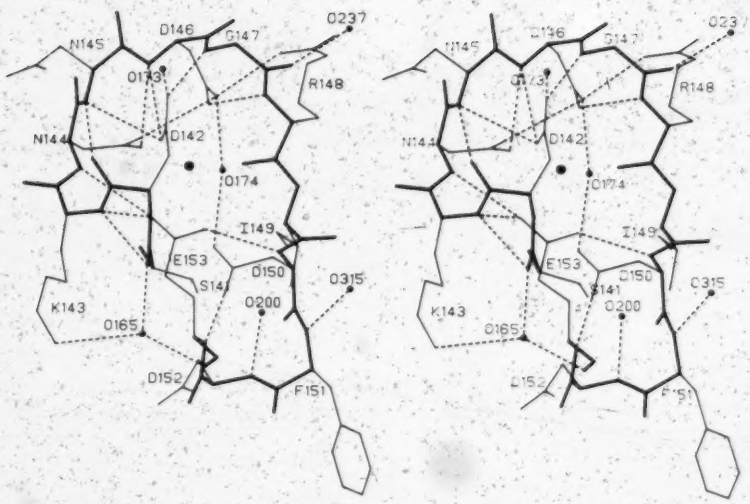
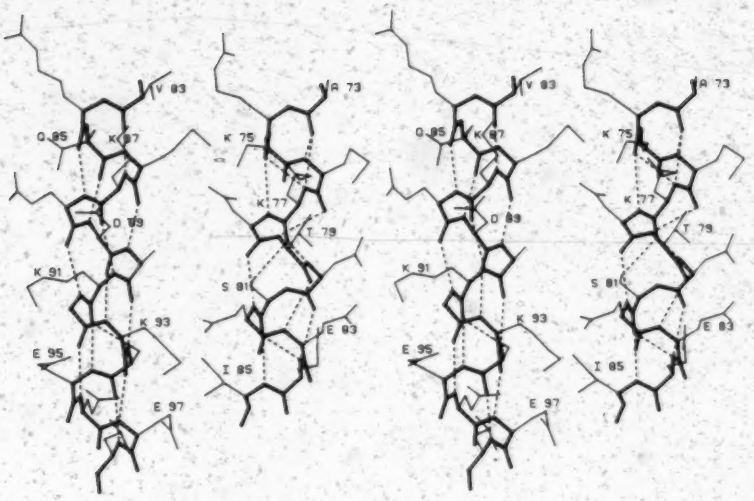
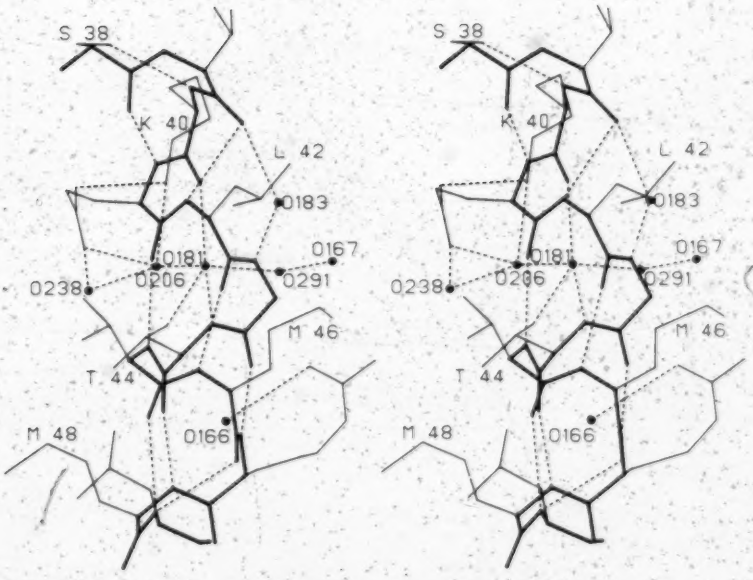


Fig. 3.5. A stereoscopic view of Ca^{2+}-binding loop IV in turkey skeletal TnC. The backbone is represented by a thick line; the side chains by a thin line. The hydrogen bonding pattern shown (dashed lines) is preserved amongst the standard Ca^{2+}-binding loops of the HLH family. Conserved water molecules are depicted as small filled circles, the Ca^{2+} ion by a larger filled circle. The β-sheet hydrogen-bonding interactions of Ile149 are not shown.

a


V 83
Q 85
K 87
D 89
K 91
K 93
E 95
E 97
A 73
K 75
K 77
T 79
S 81
E 83
I 85

b


S 38
K 40
L 42
O183
O181
O167
O238
O206
O291
M 46
T 44
M 48
O166

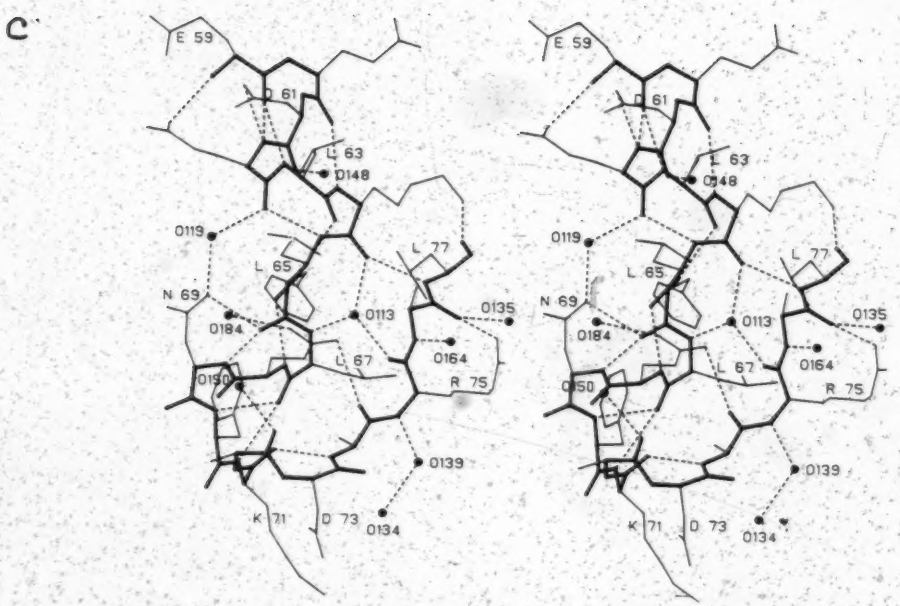
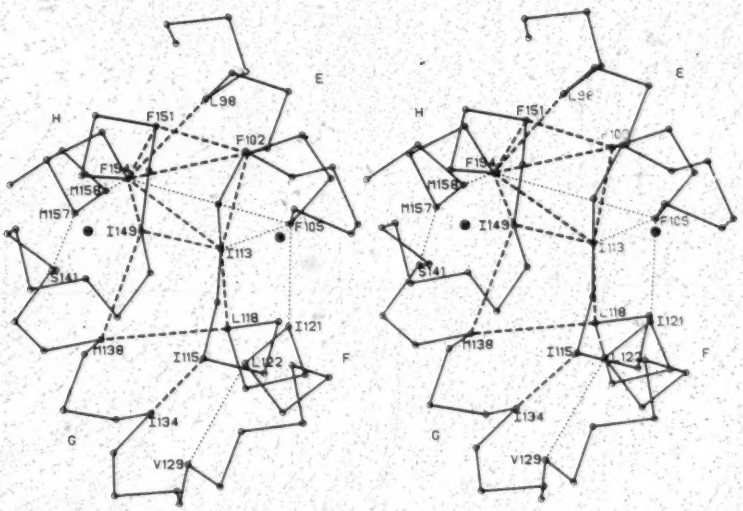
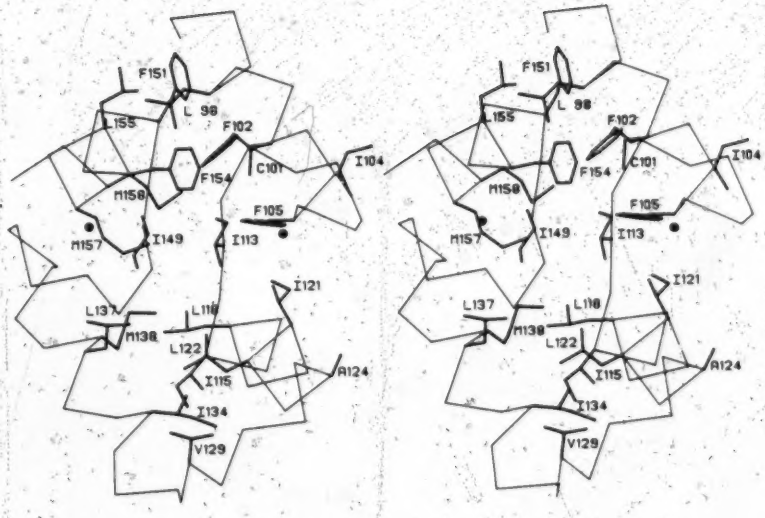


Fig. 3.6. (a) A stereoview of the two central helices in TnC (left) and CaM (right). In this and in parts (b) and (c) the main chain is shown in thick lines, the side chains in thin lines and the hydrogen bonds are dashed. (b) A stereoview of the kink in helix B of TnC. Two water molecules (O181 and O183) are inserted into the helix at the point where it bends. (c) A view of helix D in parvalbumin. The hydrogen bonding in this helix is predominently that of a 3_{10} helix. Water O113 forms hydrogen bonds to backbone atoms of the helix.

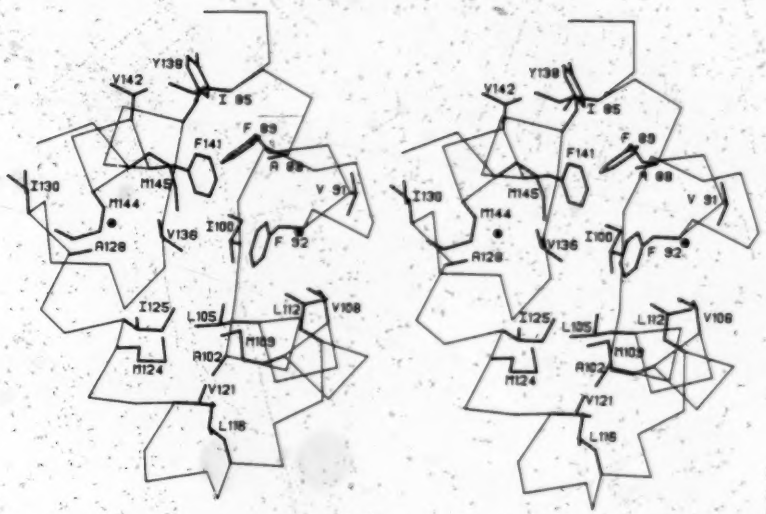
a


E
H
F151
L98
F102
F154
M158
M157
I149
F105
I113
S141
L118
I121
M138
I115
L122
F
I134
G
V129

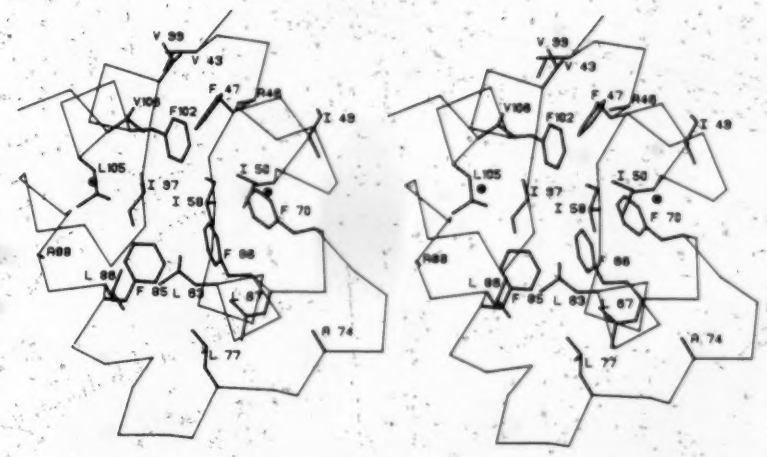
b


F151
L 98
L155
F102
C101
I104
F154
M158
F105
M157
I149
I113
I121
L137
L118
M138
L122
I115
A124
I134
V129

c


Y139
V142
I 95
F 89
F141
A 88
M145
V 91
I130
M144
V136
I100
F 92
A128
I125
L105
L112
V108
M124
A102
M103
V121
L118

d


V 99
V 43
F 47
R48
I 49
V108
F102
L105
I 97
I 58
F 70
R88
L 96
F 66
F 95
L 63
L 67
R 74
L 77

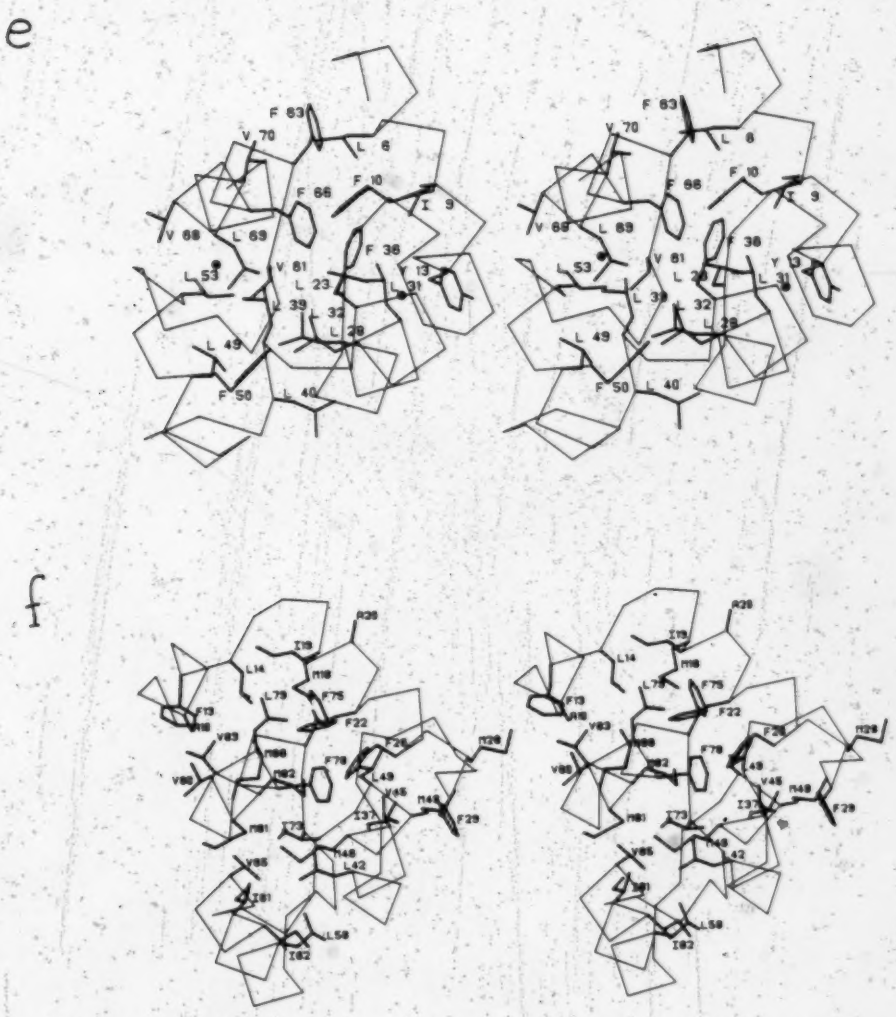


Fig. 3.7. Stereoscopic views of the disposition of hydrophobic residues in the HLH proteins. (a) A representation of hydrophobic contacts common to all of the Ca^{2+}-binding domains. For convenience, the C^{α}-backbone of the C-terminal domain of TnC is displayed. The analogous residues in the Ca^{2+} binding domains of the other proteins can be gleaned from Table 3.2. Our interpretation of the conserved hydrophobic contacts differs in substance from that of Sekharudu and Sundaralingam (173). Thick dashed lines between C^{α}-atoms represent hydrophobic contacts between HLH motifs that form the core of all 4 proteins; thin dotted lines represent the common contacts among 'surface' hydrophobic residues. In this view and part (b), Val161 and Gln162 have been omitted. In parts (b) to (f), in addition to the C^{α}-backbone, the complete side-chains for the hydrophobic residues are indicated by thick lines. (b) C-terminal domain of TnC. (c) C-terminal domain of CaM. (d) Parv (the N-terminal 41 amino acids which fold over the hydrophobic cup have been omitted for clarity (c.f. Figure 3.2c). (e) ICaBP. (f) N-terminal domain of TnC (the first 5 residues of the N-helix are omitted).

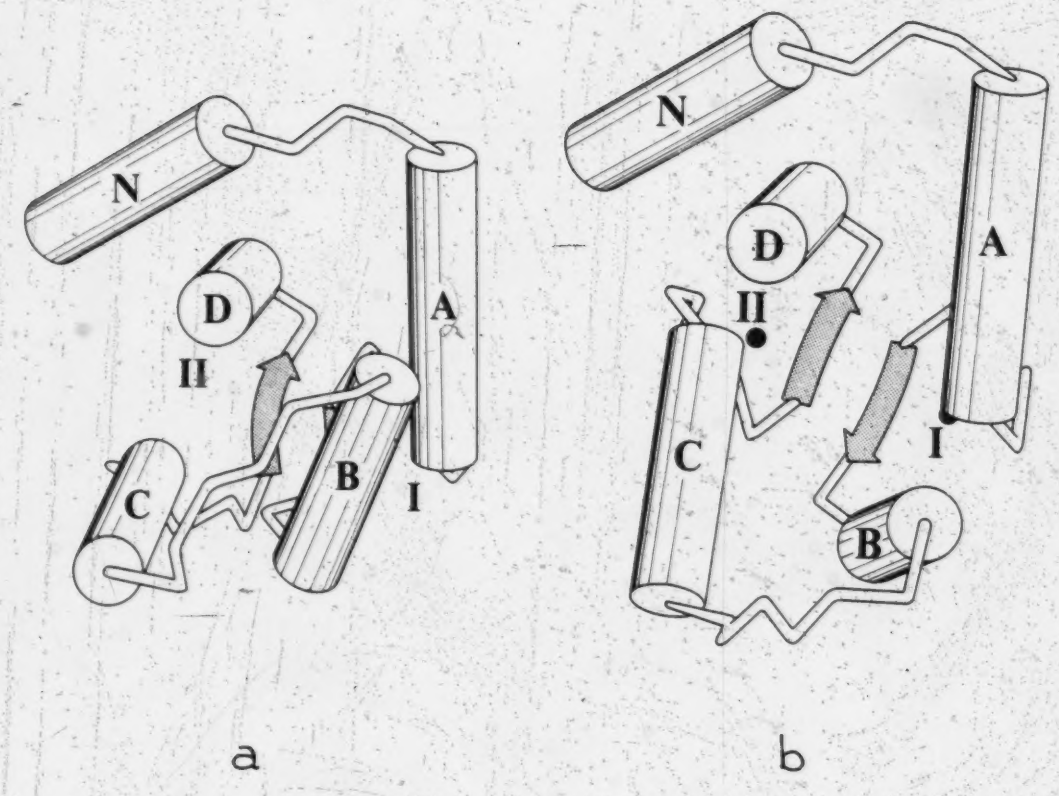


Fig. 3.8. Diagrammatic representation of the proposed Ca^{2+}-induced conformational change in the N-terminal domain of TnC (48). In this model helices N, A and D retain their relative dispositions. Helices B and C and the linker peptide move by up to 14 Å when Ca^{2+} binds. The relative dispositions of helices B and C also remain constant. (a) Ca^{2+}-free conformation of the N-terminal domain of TnC. (b) Proposed Ca^{2+}-bound conformation of this domain.

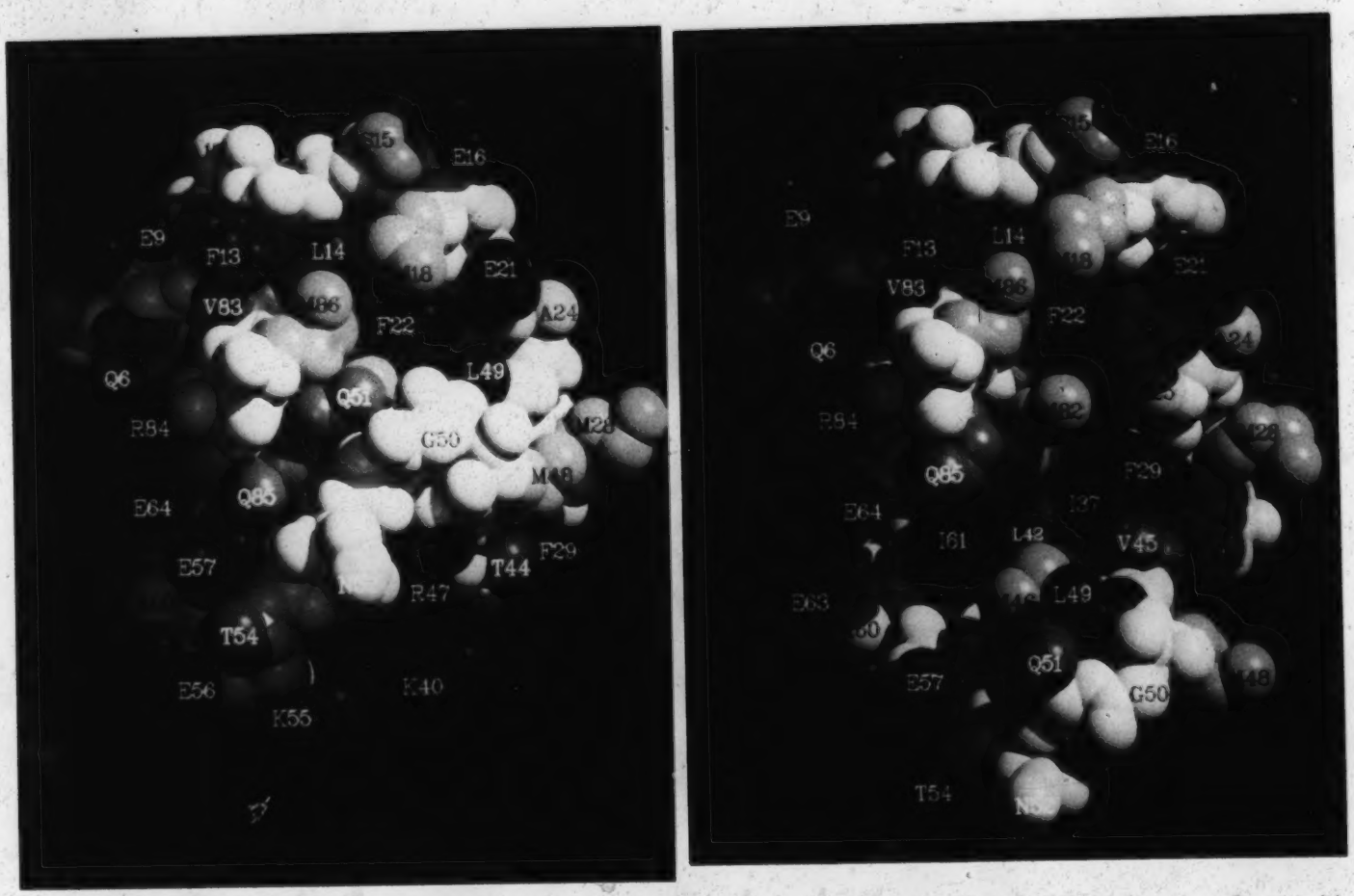


Fig. 3.9. Surface representation of the two conformational extremes of the N-terminal domain of TnC. Atoms are shown with their appropriate van der Waals radii. These views correspond to the two views in Figure 3.8. This computer graphics rendering was produced with software (RASTER3D) developed by David Bacon. The color coding for the side chains is: shades of green, aliphatic residues; brown, aromatic residues; blue are basic amino-acids, red are acidic, oranges and pinks are hydrophilic, yellow represents methionine. (a) Ca^{2+}-free conformation. (b) Proposed Ca^{2+}-bound conformation.

CHAPTER 4

TWO TRIFLUOPERAZINE BINDING SITES ON CALMODULIN PREDICTED FROM COMPARATIVE MOLECULAR MODELLING WITH TROPONIN-C[1]

INTRODUCTION

Calmodulin (CaM) is an ubiquitous, small, acidic protein of molecular weight 16,700 (1). It has four calcium-binding sites, two in the amino-terminal (N-terminal) domain with $K_d \sim 10^{-5}$ M and two in the carboxy-terminal (C-terminal) domain of slightly higher Ca^{2+} binding affinity (2,3,4). CaM plays a pivotal role in many Ca^{2+}-dependent intracellular functions including the regulation of the reactions of a number of enzymes: phosphodiesterase (5,6), myosin light chain kinase (8,9,10), erythrocyte Ca^{2+}-ATPase (11,12), brain adenylate cyclase (13,14,15), phosphorylase kinase (16) and nicotinamide dinucleotide kinase (7).

The Ca^{2+}-saturated form of CaM is the active form. The enhancement of enzymatic activity in all these systems is thought to be due to conformational changes in the enzymes induced upon binding the four-calcium ($4Ca^{2+}$) form of CaM. Therefore, Ca^{2+} efflux/influx of the cell controls CaM activation which in turn modulates these intracellular processes. Additionally, CaM-mediated activation of these enzymes can be inhibited by various drugs including those of the phenothiazine family (17,18).

Structurally, CaM belongs to the family of calcium-binding proteins including troponin-C (TnC), parvalbumin and the intestinal Ca^{2+}-binding protein. These proteins exhibit a common structural motif of helix-loop-helix frequently referred to as the EF-hand (19). The EF-hand comprises approximately thirty-three residues in a linear sequence. The central Ca^{2+}-binding loop has 12 residues and the two flanking α-helices have from 10 to 14 residues. The angle between the helical axes, E-F, varies from approximately 110° in

---

[1] A version of this chapter has been published. Strynadka & James 1988. *Proteins: Structure, Function, and Genetics* 3:1-17. Relevant material published subsequent to the paper is given in Chapter 7.

the calcium-filled conformation to between 133° to 150° in the calcium-free conformation (20). In the Ca^{2+}-filled conformation, there is a very strong conservation of tertiary structure of the binding loops among the several proteins whose crystal structures have been refined (21).

The structure of TnC from turkey skeletal muscle has been refined to 2.0 Å resolution (22). The present crystallographic agreement factor, $R(=\Sigma||F_o|-|F_c||/\Sigma|F_o|$, where $|F_o|$ and $|F_c|$ are the observed and calculated structure factor amplitudes) is 0.155. The root mean square deviations of the refined structure from standard amino-acid geometry are small. The current TnC structure consists of 160 amino acids (there is no interpretable electron density for the first residue and the last side chain of TnC), two Ca^{2+} ions in the high affinity sites of the Ca^{2+}/Mg^{2+} C-terminal domain and 156 ordered solvent molecules refined as oxygen atoms. The N-terminal domain has no Ca^{2+} ions bound and the helices have markedly different interaxial angles to the equivalent helices of the C-terminal domain (20,23,24). It has been proposed that the crystal structure of TnC represents the conformation of TnC in the relaxed state of muscle (24). Thus, on release of Ca^{2+} from the sarcoplasmic reticulum, the N-terminal domain of TnC binds two calcium ions, triggering the contraction event (23). It is the 4-Ca^{2+} form of TnC that we have used as a template to construct the model of CaM.

Not only does CaM have a high degree of sequence identity with TnC (51%) they also share many similar physicochemical properties (25,26). In fact, bovine brain CaM can substitute effectively for TnC in activating the actomyosin ATPase system of skeletal muscle (27,28). Therefore, the refined crystal structure of skeletal TnC should provide an excellent basis for the construction of a model for CaM.

A preliminary report of the crystal structure of CaM (29) has confirmed that indeed it has a molecular architecture closely resembling that of TnC (20,24). However, at present, neither the details of the individual side chain conformations nor of the calcium binding loop structures are available in published or accessible atomic coordinates.

Therefore we have built a complete atomic model of CaM in order to provide further structural data for the interpretation of the many physical measurements for that molecule in solution and to evaluate our current model-building procedures. Two previous model-building attempts on the CaM structure have been published but these were based upon the structures of carp parvalbumin and bovine intestinal Ca^{2+} binding protein (30,31). The description of the present model (and ultimately the comparison of our model with the refined structure of CaM) may reveal a more faithful prediction of the CaM structure. Our modelling has not been influenced in any way by prior knowledge of the CaM atomic coordinates. We discuss the validity of the modelling in terms of the known experimental data on CaM function.

METHODS

The amino acid sequences of bovine brain CaM and chicken skeletal TnC have been aligned as shown in Figure 1. The sequence of turkey skeletal TnC is not available but amino-acid composition analysis and the electron density from the X-ray crystallographic analysis indicate only 2 amino acid differences. For one of these in the turkey TnC a glutamate replaces an alanine at position 99.

In order to construct a CaM model the sequence alignment of Fig. 4.1 and the computer program MUTATE (R.J. Read, unpublished) were used to replace the side chains of the $4Ca^{2+}$-TnC model (23) with the homologous side chains of CaM. This replacement was done by retaining atoms common to the native and mutated residues and then building any additional atoms required in the standard conformation for that amino acid residue. The two most significant structural changes are the loss of the first ten residues of the N-terminal helix and a three residue Lys-Gly-Lys deletion from the D/E interdomain helical linker of TnC (see Fig. 4.1). This internal deletion was compensated by calculating a rotational matrix and a translation vector to accomplish the joining of the two CaM residues directly preceding and following the deletion (Asp80, Ser81). In TnC, the C-terminal residues Gly160, Val161 and Gln162 depart from an α-helical conformation

due to an intermolecular contact in the crystal structure. In our model of CaM, the C terminal three residues (Thr146, Ala147, Lys148) of helix H (see Fig. 4.1) were given an α-helical conformation.

In order to relieve unacceptably close van der Waals contacts and to correct for the inappropriate geometry obtained when mutating a glutamic acid to a proline residue at position 66, 750 cycles of conjugate-gradient energy minimization were carried out on the modelled CaM. The energy minimization was done using the GROMOS library of computer programs (32) The potential function used in this suite of programs is essentially that described by Karplus and Van Gunsteren (33). The potential function parameters were those of set 37 D of GROMOS with an 8 Å cutoff. All bond lengths and angles were optimized to fit the potential used in the energy minimization. Electrostatic charges including the contributions from calcium ions, were not considered in the calculations. Calcium-ligand binding distances within loops I, II, III, IV were idealized using a suite of programs on a Silicon Graphics (Iris) workstation.

A similar approach as outlined above was used to construct a model of the calcium free form of calmodulin. The calcium free conformation of the N terminal domain of the highly refined 2 calcium turkey TnC structure was used as a template to model a calcium-free N terminus of calmodulin, again using the sequence alignment of Fig. 4.1 and replacing side chains with the MUTATE program. In turn, the calcium free conformation of the N terminal domain was used as a template to model a calcium free C terminal domain, with residues of helixes E, F, G and H substituting into the analogous site in helices A, B, C and D, respectively. The appropriate transformation was used to rejoin the domains to yield a calcium-free CaM model. Finally, 1000 steps of conjugate gradient energy minimization were carried out on this CaM ($0Ca^{2+}$) model.

RESULTS AND DISCUSSION

It is clear from the amino-acid sequence alignment of Fig. 4.1 that there is extensive homology among these several Ca^{2+} sensitive regulatory proteins. The alignment of the

sequences is facilitated by the obvious sequences of the EF hands (19). It is also clear that the AB hand has high analogy with the EF hand. The two major differences between TnC and CaM, the extra 10 N-terminal residues in TnC and the deletion of the Lys Gly Lys tripeptide in the D/E helical linker of CaM can be easily accommodated. The extensive homology bodes well for successful comparative molecular modelling (34,35). Unlike some regions within the serine proteinases (35) there are no segments in which the TnC and CaM alignment are in doubt.

The extensive homology was evident even in the initial model obtained after replacing the TnC side chains with the aligned side chains of CaM. There were very few too short interatomic contacts even before the energy minimization. The only serious mainchain stereochemical problem was at position 66 in CaM in which a proline replaced a glutamate in TnC. Energy minimization compensated and corrected the stereochemistry at Pro66 without seriously altering the loop II conformation (root mean square difference in coordinates for the main chain atoms of the 12 residues of the loop before and after energy minimization was .10 Å). That proline 66 would be a conservative change was predicted much earlier by Kretsinger (19). It is straightforward to accommodate a proline at the N-terminus of a helix. The overall root mean square change in coordinates as a result of energy minimization was 0.35 Å with the largest movements associated with residues Pro66 and His107. The total drop in potential energy was from 1.05×10^4 kJ/mol to -6.9×10^3 kJ/mol. This energy does not include any electrostatic contributions from charged side chains nor does it include a contribution from the Ca^{2+} ions that were positioned to superimpose closely with the equivalent observed Ca^{2+} ion positions in the C-terminal domain of TnC. The protein ligands for each of the Ca^{2+} ions were adjusted using the interactive computer graphics package MMS (S. Dempsey) for the Silicon Graphics IRIS 3030.

Caveats Concerning this Predicted Model of CaM

Before discussing the predicted structure of CaM and detailing how we feel this

model provides molecular explanations for many of the solution properties of CaM, we would like to outline briefly how our model may differ from the refined structure of CaM in two regions.

Firstly, the polypeptide segment of the 5 residues that precede helix A could have a very different conformation. We have retained the conformation of those homologous residues in TnC primarily because there is a suitable hydrophobic site to accommodate the side chain of Leu4 of CaM in a position similar to that of Leu14 in TnC. From the preliminary publication on the crystal structure of CaM (29) this segment of chain Ala1-Thr5 looks different from that of our model.

The second expected major difference between the structure that we have predicted and that from the refined crystal structure is in the interhelical angles of the helix-loop-helix motifs. The solution of the structure of TnC showed that this is a region of great conformational variability because differences of up to 40° exist between the Ca^{2+}-free conformation and the Ca^{2+}-bound conformation (20,23). Thus, in our model of CaM, the interhelical angles for those helices flanking Ca^{2+}-binding sites I and III (A B and E F) are 107° and for those helices flanking Ca^{2+}-binding sites II and IV (C D and G H) the angle is 110°. These are, of course, the values for the interhelical angles in the $4Ca^{2+}$ model of TnC on which our structure is based (23,24). These angles differ from those reported in the initial CaM structural paper which had values for the interhelical angles of 92° (A B), 96° (C D), 97° (E F) and 107° (G H). These angles could change upon refinement of the structure but in any case it means that the N-domain and the C-domain in the crystal structure are in even more open conformations than our model would suggest. These differences in interhelical angles between the Ca^{2+}-bound domain of TnC and the two domains of CaM, if they remain after refinement of the two structures, lends credence to the proposed conformational flexibility in this part of these regulatory proteins as suggested previously (23). Specifically, we have proposed that in the N-terminal domain of TnC, interhelical flexibility may occur with the strong ion pair between Glu64 and Arg84 as a

fulcrum as it is formed in both Ca^{2+} free and the Ca^{2+}-bound conformation of this domain (23,24). The present model of CaM also has this homologous ion pair (Glu54 to Arg74) in the N-terminal domain.

Description of the General Structure of CaM

(A) CA^{2+}-SATURATED FORM Figure 2(a) is a stereo-representation of the α-carbon model of the Ca^{2+} saturated form of CaM. The corresponding α-carbon representation of the $4Ca^{2+}$-form of TnC which formed the starting template for $4Ca^{2+}$ CaM is shown in Fig. 4.2(b). The complete tertiary structure of our modelled CaM is represented in Fig. 4.3. The deleted N-terminal 10 residues in CaM is obvious from a comparison of these figures. The overall length of this model of CaM is approximately 62 Å long. The molecule is about 5 Å shorter than TnC due to the three residue deletion in the central D/E helical linker (see Fig. 4.1). This deletion has another effect that is obvious upon detailed comparison of Fig. 4.2a and 4.2b; the N-terminal domain is oriented approximately 180° relative to the C-terminal domain in CaM, whereas these domains are at approximately 120° to one another in TnC. These structural differences may provide clues to the differences in physiological response exhibited by the two proteins.

The N-terminal domain of CaM consists of four helices, A and B flank Ca^{2+} binding loop I, C and D flank Ca^{2+} binding loop II (see Fig. 4.2a). The D helix runs continuously to join the E helix through the D/E helical linker forming a long 28 residue, central helix linking the 2 domains. In the C-terminal domain helices E and F flank binding loop III with the last Ca^{2+}-binding "hand" comprising helix G, loop IV and helix H. With the interhelical angles of ~ 110° as discussed above, the Ca^{2+}-bound form of CaM is the open conformation of each domain. It is the Ca^{2+} saturated form of CaM that exhibits the regulatory control on the several enzyme systems studied.

(B) CA^{2+}-FREE FORM Figure 4 is an α carbon representation of calcium free calmodulin. As with the $4Ca^{2+}$ form of CaM, the calcium free model showed surprisingly

few bad contacts or stereochemical conflicts even before energy minimization. The overall root mean square deviation between the initial calcium free model and that after energy minimization was 0.46 Å and the total energy drop was from 1.50×10^6 to -7.7×10^3 kJ/mol.

The most obvious structural changes which occur in the $4Ca^{2+}$ to $0Ca^{2+}$ transition in the N terminal domain are as follows: Interhelical angles of the A/B, C/D helical pairs increase from 107,110° to 133,151° respectively, similar, of course, to the TnC Ca^{2+}-free N terminal domain on which it is based. The general effect of increasing the interhelical angles is to result in a more "closed" conformation of the domain (23), with residues at the N-terminal portion of helix D becoming significantly more buried. Analogous conformational changes in the C-terminal domain cause the burial of residues at the C-terminal portion of the E helix. The overall length of the calcium free model of CaM is approximately 58 Å, 4 Å shorter than the $4Ca^{2+}$ model. These values agree well with the results from solution X-ray scattering studies on CaM (J. Trewhella, personal communication).

Analysis of the $4Ca^{2+}$ and $0Ca^{2+}$ molecular models by the algorithm of Kabsch and Sanders (36) indicates that there is very little change in secondary structure between the two forms. The most significant difference is a decrease in α-helicity in the calcium-free form of less than 9%. Our model would suggest then, that the activation of CaM by the binding of Ca^{2+} is not caused by a large change in α-helicity as much of the earlier CD, UV and fluorescence work concluded (37-39), but rather by a reorientation of α-helical dipole moments already existing in the Ca^{2+}-free form. Interestingly, results from a recent vacuum-UV circular dichroism study found that the secondary structure of CaM depended only on ionic strength and was independent of the nature of the cation being used in the system. This implied to the authors that the activation of calmodulin by calcium was due to a structural reorientation rather than to a major change in secondary structure (40).

Description of the Ca^{2+} Binding Loops

Functionally, the most important regions of CaM are the four Ca^{2+} binding loops

(I, II, III, IV) of twelve residues each, that the helical pairs A/B, C/D, E/F, and G/H support. An earlier study has shown a very high level of structural conservation in the ligand binding loops of Ca^{++}-binding proteins (21). We believe this work, coupled with the relatively high sequence homology within the binding loops of TnC and CaM implies a high level of structural validity in this area of our CaM model. A view of binding sites I and II is given in Fig. 4.5a and binding sites III and IV in Fig. 4.5b.

In our model, six coordinating residues in each binding loop (marked by an asterisk in Fig. 4.1) are suitably oriented to provide oxygen atoms for binding Ca^{2+} ions. Except for the amino acid at position -x, the average Ca^{2+} to oxygen ligand distance is 2.5 Å. In this fifth coordinating position of each loop (Thr28, Asp64, Ser101, Asn137) larger distances of 4.5-4.9 Å from the side chain oxygen atom to the Ca^{2+} ion suggest that the interaction should be mediated *via* a bridging H_2O molecule as is observed in loops III and IV of TnC. This suggestion for CaM has been made previously, based on a comparison of loop conformations in Ca^{++}-binding proteins (21).

Each of the pairs of calcium binding loops are connected and stabilized by hydrogen bonding in a short antiparallel β sheet conformation. In the N terminal domain, Ile27 of loop I and Ile63 of loop II share two hydrogen bonds (2.7 Å, 2.8 Å) and similarily in the C-terminal domain, Ile100 of loop III and Val136 of loop IV share two hydrogen bonds (2.8 Å, each). This result had been deduced from earlier 2-D and NOE N.M.R. studies (41). Additional loop stabilization arises from a series of Asx (side chain oxygen of residue n to main chain nitrogen of residue n+2) and reverse turns as detailed for loops III and IV in TnC (21,24).

Factors Influencing the Ca^{2+} Binding Affinity

From the work done to date, it appears that the binding of Ca^{2+} to CaM takes place in two stages (2,3,4,41,42,43). It binds first to sites III and IV in the C-terminal domain; the second stage involves binding Ca^{2+} to sites I and II in the N-terminal domain. It is not yet clear exactly what determines the relative Ca^{2+} affinity of the individual binding loops

or whether the loops act in a cooperative manner.

Undoubtedly, there are many factors that determine the binding affinity of Ca^{2+}, not just the nature of the ligands that coordinate directly to the Ca^{2+} ion. It is likely that long range interactions, electrostatic, van der Waals and other weak non-covalent interactions will also have an effect on the ease and strength of binding Ca^{2+}. The two different conformations of CaM, the Ca^{2+}-free (closed) and the Ca^{2+}-saturated (open) forms allow us to examine the relative contributions of some of these long range interactions on Ca^{2+} binding affinity and allow a comparison of the N-terminal and C-terminal domains of CaM in this regard. Since the measured Ca^{2+} affinities of the two domains are quite similar in CaM, we should expect the weak noncovalent interactions in the two domains also to be similar.

Table 4.1 presents the solvent accessible surface areas, the number of main-chain hydrogen bonds and the number of potential ion pairs for each of the two domains in both of the Ca^{2+}-filled (open) and Ca^{2+}-free (closed) conformations. In going from the open to the closed conformation the total accessible surface area of the N-terminal domain decreases by 10%. The favorable burial of hydrophobic residues accounts for 270 $Å^2$ of the decrease. In the C-terminal domain, there is a corresponding decrease of 11% in the total accessible surface area in going from the open to the closed conformation; 260 $Å^2$ is due to the favorable burial of hydrophobic residues. Thus, it would appear that the closed conformation of the domains is to be preferred. The number of main-chain NH····O hydrogen bonds in both domains does not change significantly in going from the Ca^{2+}-bound (open) to the Ca^{2+} free (closed) conformation. Thus, our model predicts little change in secondary structure in these two states of the protein, as suggested previously.

Ion pair interactions may be important for protein stability and it is in the number of potential ion pair interactions in CaM that we see differences between the N-terminal domain and the C-terminal domain. There are also differences between the open and closed conformations of the two domains (Table 4.1). In the N-terminal domain of the $4Ca^{2+}$

CaM model, the potential ion pairs include Glu54 (on helix C) and Arg74 (on helix). This ion pair is homologous to Glu64····Arg84 in TnC and may act as the fulcrum for interhelix flexibility (23). Another conserved ion pair is Arg37 to Glu45 (see Table 4.1, Figs. 4.1 and 4.6a). In going to the Ca^{2+}-free (closed) form, there are 3 additional potential ion pair interactions possible, due to the conformational change of helices B and C (Fig. 4.6b). One of the new ones is from Glu47 to Lys75. In the closed Ca^{2+}-free conformation the change brings Glu47 into close proximity to Lys75 whereas in the open Ca^{2+}-filled form these residues are 9 Å apart. Of the Ca^{2+} sensitive regulatory proteins only CaM has a lysine at position 75 (Fig. 4.1).

In the C-terminal domain in the open Ca^{2+}-filled conformation there is only one potential ion pair (Table 4.1, Figs. 4.1 and 4.7a). Two additional ion pair interactions are possible in the closed Ca^{2+}-free conformation (Fig. 4.7b)]. Analogous to the Glu47-Lys75 ion pair of the N-terminal domain, Glu120 on helix G moves to interact potentially with Lys148 on helix H (Fig. 4.7b). Only CaM has this basic residue at position 148.

Overall, the non-covalent interactions in the N-terminal and C-terminal domains are quite similar. The extra potential ion-pair interactions in the N-terminal domain that may stabilize the Ca^{2+}-free (closed) conformation may be correlated with the observation that Ca^{2+} is firstly and preferentially bound to the C-terminal domain in CaM. Site directed mutagenesis of the CaM gene at these residues providing the ion pair interactions may provide an indication to the importance of the salt bridges in modifying Ca^{2+} affinity.

Description of the Hydrophobic Patches

The binding and release of calcium to the calcium binding domains of CaM is known to be the regulating effector of the activity of the various intracellular enzymes with which CaM interacts (5-16). The results of some studies suggest that many of these enzymes may be binding to a hydrophobic region found in each of the amino- and carboxy-terminal domains (52,69,71,72). Indeed in each of the domains of our Ca^{2+}-saturated CaM model there is a very pronounced hydrophobic patch. The hydrophobic clefts of the

N-domain and C-domain face towards each other on opposite sides of the central helix (see Fig. 4.2a and Fig. 4.3). In each case a number of hydrophobic residues are clustered around a central core composed of the highly conserved loop-stabilizing β-sheet hydrogen bonding pairs Ile27/Ile63 (N-terminal domain) and Ile100/Val136 (C-terminal domain). Table 4.2 gives a list of hydrophobic residues in the N-terminal and C-terminal domains along with their relative solvent accessibility values as calculated by the algorithm of Kabsch and Sanders (36). A view of the N-terminal and C-terminal hydrophobic regions are given in Fig. 4.6a and Fig. 4.7a, respectively. Along the entrance to both apolar cavities are a number of charged acidic residues. In the N-terminal domain these include Glu11, Glu14, Glu 47, Asp50, Glu54 and Asp78, and in the C domain these acidic residues include Glu84-87, Glu114, Glu119, Glu120, Glu123, and Glu127. Both the N and C terminal hydrophobic patches have a lysine positioned across the opening of the pocket. Lys75 lies across the bottom of the N terminal pocket and Lys148 extends across and about half way up the entrance to the C terminal hydrophobic pocket (Figs. 4.6a and 4.7a).

Model of Trifluoperazine Binding to CaM

Since none of the three dimensional structures of the enzymes that CaM activates are yet known, it seemed premature to model an exact binding site for these enzymes on the hydrophobic patches of CaM. However, it did seem plausible to fit the small, lipophilic drugs which are thought to bind in the apolar pockets and subsequently to inhibit CaM mediated activation of the various intracellular enzymes mentioned previously. These drugs include the family of antipsychotics called the phenothiazines, most studied of which is trifluoperazine (TFP). Earlier solution work has shown that a CaM molecule has 2 Ca^{2+}-dependent high affinity ($K_d \sim 5\ \mu M$) sites for TFP (44). It was suggested that one of the two high affinity binding sites was located on the N-terminal portion of the molecule and the other on the C-terminal domain of CaM. The affinity of the C-terminal site is at least one order of magnitude greater than that of the N terminal site (41). It is well documented

that the affinity of various phenothiazines for CaM is related to their hydrophobicity (45,46,47,48,49). Furthermore, studies have shown that a positive charge a certain distance from the phenothiazine ring is crucial for binding to CaM implying that electrostatic as well as hydrophobic interactions are involved (46). Ostensibly then, in the binding of TFP to CaM, one would expect the hydrophobic phenothiazine rings (see Fig. 4.8a) to stack against the residues of the hydrophobic patch, with the positively charged nitrogen of the piperazine moiety extending out of the patch to interact with the negatively charged residues framing the outer rim of the hydrophobic pocket. Additionally, several acetylation studies have shown that K75 and K148 may be involved at or proximal to the TFP binding site (50). Recent work using 10-(3-propionyloxysuccinimide)-2-(trifluoromethyl)phenothiazine (POS-TP, see Fig. 4.8b) showed that Lys148 was labelled at low molar ratios of reagent and Lys75, and Lys77 at higher molar ratios of reagent (51). This work concluded that binding of POS-TP in the C-terminal domain involved a hydrophobic patch including residues Phe89, Phe92, Phe141, Tyr138, Ile85, Ile25, Val136, Val142 and Met109, Met124, Met144 and Met145. Only a few of these residues agree with our proposed binding site in the C-terminal domain (see below and Table 4.2).

Considering the above observations and by calculating a van der Waals surface portrait of both TFP and CaM ($4Ca^{2+}$) we propose the following TFP binding sites on CaM. In the N-terminal domain the phenothiazine rings of TFP fit nicely into a groove formed by the side chains of Val35, Leu39, Met36, Leu18, Met71, Met72, Phe19 and Ala15 (see Figs. 4.9a and 4.10a). The positively charged nitrogen of the piperazine ring can then easily interact with one or both of the acidic side chains of Glu14 and Glu11 which extend off helix A directly outside the hydrophobic patch. Lys75 runs along side the trifluoro group at position 2 of the phenothiazine ring.

In the C-terminal domain hydrophobic patch, the binding site of the phenothiazine ring includes residues Val91, Val108, Leu112, Met109, Phe92, Met144, Met145 and Ala88. In our model Tyr138, Phe89, Phe141, Ile85, Ile125, Val142 and Met124 are too

far from the surface of the hydrophobic pocket to interact with TFP. The positively charged nitrogen of the piperazine ring could easily interact with Glu84 and or Glu87 which extend off the E helix. Lys148 again like Lys75, runs along the side of the trifluoro moiety. Figs. 4.9b and 4.10b show a view of TFP bound in the C-terminal domain. Fig. 4.1 shows the relative orientation of the 2 TFP molecules on the whole CaM. Near both TFP binding sites there is a small cavity formed near the interface of Met71 and Met72 in the N terminus and Met144 and Met145 in the C terminus. These cavities could accommodate a propyl or similar length group on position 3 of the phenothiazine ring. It would be of interest to see if such an analog would invoke tighter binding. From Figs. 4.9a and 4.10a, one can easily see how an acetylating reagent like POS-TP which has a neutrally charged acetylating succinimide group at the same position as the positively charged piperazine ring in TFP (see Fig. 4.8a,b) could flip over to acetylate the adjacent lysine (K75 or K148). It is also clear from these figures that Lys148 is much less sterically hindered than Lys75 and therefore should be more chemically reactive with POS-TP as was observed. Accessibility values (36) for Lys75 and Lys148 in the $4Ca^{2+}$ CaM model are 117 and 225 respectively.

Several studies have indicated that the greater affinity that phenothiazine has for CaM in the presence of Ca^{2+} is the result of modified exposure of hydrophobic patches in the Ca^{2+} bound conformation (44,51,52,53,54,55). Comparison of the Ca^{2+}-bound and the Ca^{2+}-free conformation of the N-terminal domain of CaM seems to bear these observations out (see Figs 6a and 6b). The hydrophobic patch is significantly more buried in the absence of Ca^{2+}. Of the eight hydrophobic residues involved in binding TFP in the N-terminus of the Ca^{2+}-free state only three residues, Phe19, Leu18 and Ala15 are still suitably exposed to interact with the hydrophobic phenothiazine rings of TFP. Glu11 and Glu14 also retain their relative position to bind the charged piperazine moiety of TFP. Collectively these residues may explain the weak affinity binding of TFP to CaM in the absence of Ca^{2+}. Analogously in the C-terminal domain, only Phe92, Val91 and Ala88 are

sufficiently exposed to bind TFP along with Glu84 and Glu87 (see Figs. 4.7a and 4.7b). Table 4.1 shows the change in accessibility of key residues of these hydrophobic patches from the calcium filled to calcium free state of the N terminal and C terminal domains. Residues marked with an asterisk are thought to be involved in TFP binding.

A recent paper using amine-directed chemical modification reagents showed the very high, Ca^{2+}-dependent reactivity of Lys75 (56). They observed that the adjacent lysine 77, had very low reactivity and suggested that this may be due to the fact that Lys77 may be involved in an ion pair with Asp78. The very dramatic calcium dependence of Lys75 reactivity may be explained by our model in two ways. Firstly, as mentioned earlier, in the calcium filled to calcium free transition, Glu47 swings in very closely to Lys75 where they could form a salt bridge, thus tying up Lys75 and decreasing its reactivity with reagent. Secondly, and perhaps relatedly, to the accessibility value of Lys75 drops dramatically (117→12) in going from the Ca^{2+}-bound to the Ca^{2+}-free conformation.

Lys75 and Lys148 have been shown to have increased protection from acetylation when CaM is complexed with myosin light chain kinase (57) and calcineurin (58). These results coupled with the observed inhibitory effect of TFP on the activation of these enzymes suggests a common binding domain for at least some of the CaM target enzymes. Site specific mutation of Glu82, Glu83 and Glu84 to lysine residues resulted in the expression of a CaM mutant with a 70% decrease in activation of myosin light chain kinase (59). Thus, Glu84 may be involved in MLCK binding to CaM as it is for our hypothetical TFP binding site.

Binding of certain basic amphiphilic α-helical peptides to the C terminus of CaM can block its ability to activate some target enzymes including MLCK and phosphodiesterase (60,61). Tight binding of these amphiphilic helixes requires multiple positively charged residues on the peptide, so regions with a high density of negatively charged, acidic residues should attract such peptides. As mentioned earlier, such a region exists as a

cluster of glutamic acids surrounding the C terminal hydrophobic patch. Both α-helical inhibitory peptides, mellitin and mastoporan have lysine and arginine clusters at the termini of their helices and a series of hydrophobic residues predominantly along one face of the helix. Perhaps the binding of these peptides involves the interaction of their basic residues with the acidic patch on CaM and the hydrophobic moieties of the helix pointing in towards the hydrophobic patch of CaM. Such a mode of binding would resemble what we have proposed for TFP and would sterically block the binding of target enzymes to the hydrophobic patch. The Ca^{++} dependence of binding of both melletin and mastoporan would be due, as with TFP, to their exclusion from the hydrophobic patch as it becomes buried upon Ca^{++} release.

Other structural features of interest include the following: the clustering of charged acidic residues along the interdomain helix (Asp78, Asp80, Glu82, Glu83, and Glu84, Glu87, and the position of tyrosine residues. Tyr99 is directed from loop III towards the surface where it is accessible to solvent and is stacked against the side chain of Gln135 (Fig. 4.5b). Tyr138 points away from loop IV towards helix E where it comes into close proximity with Phe89 and Phe141. This is consistent with previous solution work that indicated the high pK of Tyr138 was due to its hydrophobic environment (62). The unique trimethyl-lysine at position 115 is located between helixes F and G and does not appear to be interacting with other residues. The single His107 residue on helix F is fully accessible to solvent.

CONCLUSIONS

With the ever increasing numbers of protein sequences being determined by DNA sequencing methods, there is an increasing need to predict the proteins three-dimension structure so that further experiments may be designed. Unfortunately, to date, the secondary structure prediction algorithms are unreliable (63). Until fundamental advances in the solution of the protein folding problem are forthcoming the method of comparative molecular modelling (34) provides the greatest promise of success (35). Furthermore even

with the assumption that the secondary structure is correctly predicted, the resulting profile of helices and β-strands is of limited use. Sites of functional importance (active sites on enzymes, effector sites, drug binding sites, etc.) are usually composed of amino-acids distant in the primary sequence but brought together by the tertiary folding in the final structure.

As a predictive method, the computer modelling of a protein structure from its sequence and with the knowledge of a crystallographically-defined three-dimensional structure of a highly homologous molecule should provide a more accurate and useful description of the molecule. This method is not without its limitations either (34,35,69). It is the regions surrounding those that are highly conserved in the active site that are different among different enzymes. These loops usually provide the substrate specificity and therefore provide the greatest trouble in modelling.

In order to result in as faithful a model as possible, the protein chosen for a template should have as high a sequence analogy with the unknown protein as possible. Troponin-C provides such a template for calmodulin. These proteins have 51% sequence identity and many similar physicochemical characteristics. Within the framework of the two caveats we acknowledge, the structure we describe here should be as faithful a description of calmodulin as possible. Our model building efforts are directly testable. We have deposited the coordinates with the Brookhaven Protein Data Bank so they can be compared with the coordinates derived from the refined crystal structure of calmodulin when they are available. Lastly the two proposed binding sites of TFP have given insight into the design of new inhibitors based on the phenothiazine structure that could be more specific and that could bind more tightly than existing inhibitory molecules.

REFERENCES

1. Cheung, W.Y. Calmodulin plays a pivotal role in cellular recognition. Science 207:19-27, 1980.

2. Anderson, A., Forsen, S., Thulin, E., Vogel, H.J. ^{113}Cd nuclear magnetic resonance studies of proteolytic fragments of calmodulin. Assignment of strong and weak cation binding sites. Biochemistry 22:2309-2313, 1983.

3. Aulabaugh, A., Niemczura, W.P., Gibbons, W.A. High field proton NMR studies of tryptic fragments of calmodulin. A comparison with the native protein. Biochem. Biophys. Res. Comm. 118:225-232, 1984.

4. Klevit, R.E., Dalgarno, D.C., Levine, B.A., Williams, R.J.P. The nature of the Ca^{2+}-dependent conformational change. Eur. J. Biochem. 139:109-114, 1984.

5. Lin, Y.M., Liu, Y.P., Cheung, W.Y. Cyclic 3',5'-nucleotide phosphodiesterase. Purification, characterization, and active form of the protein activator from bovine brain. J. Biol. Chem. 249:4943-4594, 1974.

6. Teo, T.S., Wang, J.H. Mechanisms of activation of a cyclic adenosine 3',5' monophosphate phosphodiesterase from bovine heart by calcium ions. J. Biol. Chem. 248:5950-5955, 1973.

7. Anderson, J.M., Cormier, M.J. Calcium-dependent regulator of NAD kinase in higher plants. Biochem. Biophys. Res. Commun. 84:595-602, 1978.

8. Hathaway, D.R., Adelstein, R.S. Human platelet myosin light chain kinase requires the calcium-binding protein calmodulin for activity. Proc. Natl. Acad. Sci. U.S.A. 76:1653-1657, 1979.

9. Dabrowska, R., Sherry, J.M.F., Avomatorio, D.K., Hartshorne, D.J. Modulator protein as a component of the myosin light chain kinase from chicken gizzard. Biochemistry 17:253-258, 1978.

10. Waisman, D.M., Singh, T.J., Wang, J.N. The modulator-dependent protein kinase. J. Biol. Chem. 253:3387-3390, 1978.

11. Lynch, T.J., Cheung, W.Y. Human erythrocyte Ca^{2+}, Mg^{2+}-ATPase: Mechanism of stimulation by calcium. Arch. Biochem. Biophys. 194:165-170, 1979.

12. Hanahan, D.J., Tavern, R.D., Flynn, D.D., Eckholm, J.E. The interaction of Ca^{2+}/Mg^{2+} ATP'ase activator protein and Ca^{2+} with human erythrocyte membranes. Biochem. Biophys. Res. Commun. 84:1009-1015, 1978.

13. Brostrom, M.A., Brostrom, C.O., Breckenridge, B.M., Wolff, D.J. Regulation of adenylate cyclase from glial tumor cells by calcium and a calcium-binding protein. J. Biol. Chem. 251:4744-4750, 1976.

14. Brostrom, M.A., Brostrom, C.O., Wolff, D.J. Calcium-dependent adenylate cyclase from rat cerebral cortex: Activation by guanine nucleotides. Arch. Biochem. Biophys. 191:341-350, 1978.

15. Westcott, K.R., Laporte, D.C., Storm, D.R. Resolution of adenylate cyclase sensitive and insensitive to Ca^{2+} and calcium-dependent regulatory protein (CDR) by CDR-sepharose affinity chromatography. Proc. Natl. Acad. Sci. U.S.A. 76:204-208, 1979.

16. Cohen, P., Burchell, A., Foulkes, J.G., Cohen, P.T.W., Vanaman, T.C., Nairn, A.C. Identification of the Ca^{++}-dependent modulator protein as the fourth subunit of rabbit skeletal phosphorylase kinase.

17. Crow, T.J., Deakin, J.W.F., Johnstone, E.C. Stereospecificity and clinical potency of neuroleptics. Nature 267:183-4, 1977.

18. Earl, C.Q., Prosialeck, W.C., Wiess, B. Inhibition of calmodulin activity by alpha and renergic antagonists. Fed. Proc. 41:1565-1569, 1982.

19. Kretzinger, R.H. Structure and evolution of calcium-modulated proteins. CRC Crit. Rev. Biochem. 8:119-174, 1980.

20. Herzberg, O., James, M.N.G. Structure of the calcium regulatory muscle protein troponin-C at 2.8 Å resolution. Nature (London) 313:653-659, 1985.

21. Herzberg, O., James, M.N.G. Common structural framework of the two Ca^{2+}/Mg^{2+} binding loops of troponin C and other Ca^{2+} binding proteins. Biochemistry 24:5298-5302, 1985.

22. Herzberg, O., James, M.N.G. Refined crystal structure of troponin C from turkey skeletal muscle at 2.0 resolution. J. Mol. Biol. (Submitted).

23. Herzberg, O., Moult, J., James, M.N.G. A model for the Ca^{2+} induced conformational transition of troponin C and other Ca^{2+} binding proteins. J. Biol. Chem. 261:2638-2644, 1986.

24. Herzberg, O., Moult, J., James, M.N.G. Calcium binding to skeletal muscle troponin C and the regulation of muscle contraction. Calcium and Cell Function (Ciba Foundation Symposium 122), Wiley, Chichester, 120-139, 1986.

25. Stevens, F.C., Walsh, M., Ho, H., Teo, J.H., Wang, J.H. Comparison of calcium-binding proteins. Bovine heart and brain protein activators of cyclic nucleotide phosphodiesterase and skeletal muscle troponin C. J. Biol. Chem. 251:4495-4500, 1976.

26. Watterson, D.M., Harrelson, W.G. Jr., Keller, P.M., Sharief, F., Vanaman, T.C. Structural similarities between the Ca2+-regulatory proteins of 3'5'-cyclic nucleotide phosphodiesterase and actomyosin ATP'ase. J. Biol. Chem. 251:4501-4513, 1976.

27. Amphlett, G.W., Vanaman, T.C., Perry, S.V. Effect of the troponin C-like protein from bovine brain (brain modulator protein 1) on the Mg^{++}-stimulated ATPase of skeletal muscle actomyosin. FEBS Lett. 72:163-168, 1976.

28. Castellani, L., Morris, E.P., O'Brien, E.J. Calmodulin as a model for troponin C. Biochem. Biophys. Res. Commun. 96:558-565, 1980.

29. Babu, Y.S., Sack, J.S., Greenhough, T.J., Bugg, C.E., Means, A.R., Cook, W.J. Three dimensional structure of calmodulin. Nature 315:37-40, 1985.

30. O'Neil, K.T., Delgrado, W.F. A predicted structure of calmodulin suggests an electrostatic basis for its function. Proc. Natl. Acad. Sci. 82:4954-4958, 1985.

31. Aulabaugh, A., Niemczura, W.P., Blundell, T.L., Gibbons, W.A. A study of the interactions between residues in the C-terminal half of calmodulin by one and two-dimensional NMR methods and computer modelling. Eur. J. Biochem. 143:409-418, 1984.

32. Van Gunsteren, W.F., Berendsen, H.J.C. Gromos Laboratory of Physical Chemistry, University of Groningen, Nijenborgh 16, 9747 AG Groningen, The Netherlands, 1984.

33. Van Gunsteren, W.F., Karplus, M. Effect of constraints on the dynamics of macromolecules. Macromolecules 15:1528-1544, 1982.

34. Greer, J. Comparative model-building of the mammalian serine proteases. J. Mol. Biol. 153:1027-1042, 1981.

35. Read, R.J., Brayer, G.D., Jurásek, L., James, M.N.G. Critical Evaluation of comparative model building of *Streptomyces griseus* trypsin. Biochemistry 23:6570-6575, 1984.

36. Kabsch, W., Sanders, C. Dictionary of protein secondary structure:pattern recognition of hydrogen bonded and geometrical features. Biopolymers 22:2577-2637, 1983.

37. Van Eerd, J.P., Kawasaki, Y. Ca^{++} induced conformational changes in the Ca^{++} binding component of troponin C. Biophys. Res. Comm. 47:859-865, 1972.

38. Murray, A.C., Kay, C.M. Hydrodynamic and optical properties of troponin C. Demonstration of a conformational change upon binding calcium ion. Biochemistry 11:2622-2627, 1972.

39. Johnson, J.D., Potter, J.D. Detection of two classes of Ca^{2+} binding sites in troponin C with circular dichroism and tyrosine fluorescence. J. Biol. Chem. 253:3775-3777, 1978.

40. Hennessey, J.D., Parthasavathy, M.W., Johnson, C.W., Malencik, D.A., Anderson, S.A., Schimerlik, M.I., Shalifin, Y. Conformational transitions of calmodulin as studied by vacuum-U.V. CD. Biopolymers 26:561-571, 1987.

41. Dalgarno, D.C., Klevit, R.E., Levine, B., Williams, R.J.P., Dobrowolski, Z. ^1H-NMR studies of calmodulin. Resonance assignments by use of tryptic fragments. Eur. J. Biochem. 138:281-289, 1987.

42. Ikura, M., Hiraoki, T., Mikichi, K., Mikun, T., Yazawa, M., Yagi, K. Nuclear magnetic resonance studies on calmodulin: Ca2+-induced conformational change. Biochemistry 22:2573-2579, 1983.

43. Ikura, M., Hiraoki, T., Hiluchi, K., Minowa, O., Yamaguchi, N., Yazawa, W., Yagi, K. Nuclear magnetic resonance studies on calmodulin: Ca^{2+} dependent spectral change of proteolytic fragments. Biochemistry 23:3121-3128, 1984.

44. Dalgarno, P.C., Klevit, R.E., Levine, B.A., Scott, G.M., Williams, R.J.P. The nature of the trifluoperazine binding sites on calmodulin and troponin-C. Biochim. Biophys. Acta 791:164-172, 1984.

45. Richman, P.G., Klee, C.B. Conformation-dependent nitration of the protein activator of cyclic adenosine 3',5'-monophosphate phosphodiesterase. Biochemistry 17:928-935, 1978.

46. Prozialeck, W.C., Weiss, B. Inhibition of calmodulin by phenothiazines and related drugs: structure-activity relationships. J. Pharmacol. Exper. Ther. 222:509-514, 1982.

47. Levin, R.M., Weiss, B. Specificity of the binding of trifluoperazine to the calcium-dependent activator of phosphodiesterase and to a series of calcium-binding proteins. Biochem. Biophys. Acta 540:197-204, 1978.

48. Levin, R.M., Weiss, B. Binding of trifluoperazine to the calcium-dependent activator of cyclic nucleotide phosphodiesterase. Mol. Pharmacol. 13:690-694, 1977.

49. Levin, R.M., Weiss, B. Selective binding of anti-psychotics and other psychoactive agents to the calcium-dependent activator of cycle nucleotide phosphodiesterase. J. Pharmacol. Exp. Ther. 208:454-459, 1979.

50. Giedroc, D.P., Sinha, S.K., Brew, K., Puett, D. Differential trace labelling of calmodulin: Investigation of binding sites and conformational states by individual lysine reactivities. J. Biol. Chem. 260:13406-13413, 1985.

51. Faust, F., Slisz, M., Jarrett, M. Calmodulin is labelled at lysine 148 by a chemically reactive phenothiazine. J. Biol. Chem. 262:1938-41, 1987.

52. Laporte, D.C., Wierman, B.M., Storm, D.R. Calcium-induced exposure of a hydrophobic surface on calmodulin. Biochemistry 19:3814-3918, 1980.

53. Tanaka, T., Hidaka, H. Hydrophobic regions function in calmodulin-enzyme(s) interactions. J. Biol. Chem. 255:11078-11080, 1980.

54. Klevit, R.E., Levine, B.A., Williams, R.J.P. A study of calmodulin and its interaction with trifluoperazine by high resolution ^{1}H NMR spectroscopy. FEBS Lett. 123:25-29, 1981.

55. Krebs, J., Carafoli, E. Influence of Ca^{2+} and trifluoperazine on the structure of calmodulin. Eur. J. Biochem. 124:619-627, 1982.

56. Mann, D., Vanaman, T.C. Chemical modification as a probe of calmodulin function. Meth. in Enzymol. 139:417-433, 1987.

57. Jackson, A.E., Carraway, K.L., Puett, D., Brew, K. Effects of the binding of myosin light chain kinase on the reactivities of calmodulin lysines. J. Biol. Chem. 261:12226-12232, 1986.

58. Manalan, A.S., Klee, C.B. Affinity selection of chemically modified proteins: Role of lysyl residues in the binding of calmodulin to calcineurin. Biochemistry 26:1382-1320, 1987.

59. Craig, T.A., Watterson, D.M., Prendergrast, F.G., Haiech, J., Roberts, D.M. Site specific mutagenesis of the α-helices of calmodulin. Biochemistry 26:1382-1390, 1987.

60. Malencik, D.A., Anderson, S.R. Binding of simple peptides, hormones and neurotransmitters by calmodulin. Biochemistry 21:3480-3486, 1982.

61. Malencik, D.A., Anderson, S.R. Peptide binding by calmodulin and its proteolytic fragments and by troponin C. Biochemistry 23:2420-2428, 1984.

62. Seamon, K.B. Calcium and magnesium dependent conformational states of calmodulin as determined by nuclear magnetic resonance. Biochemistry 19:207-215, 1980.

63. Kabsch, W., Sander, C. How good are predictions of protein secondary structure? FEBS Lett. 155:179-182, 1983.

64. Collins, J.N., Potter, J.D., Norn, J.H., Wilshire, G., Jackman, N. The amino acid sequence of rabbit skeletal muscle troponin C: gene replication and homology with calcium-binding proteins from carp and hake muscle. FEBS Lett. 36:268-272, 1973.

65. Wilkinson, J.M. The amino acid sequence of troponin C from chicken skeletal muscle. FEBS Lett. 70:254-256, 1976.

66. Van Eerd, J.P., Takahashi, K. Determination of the complete amino acid sequence of bovine cardiac troponin C. Biochemistry 15:1171-1180, 1975.

67. Watterson, D.M., Sharief, F., Vanaman, T.C. The complete amino acid sequence of the Ca^{2+}-dependent modulator protein (calmodulin) of bovine brain. J. Biol. Chem. 255:962-975, 1980.

68. McDowell, J.J.H. Trifluoperazine hydrochloride, a phenothiazine derivative. Acta Cryst. B36:2178-2181, 1980.

69. Jarret, H. The synthesis and reaction of a specific affinity label for the hydrophobic drug-binding domains of calmodulin. J. Biol. Chem. 259:10136-10144, 1984.

70. Moult, J., James, M.N.G. An algorithm for determining the conformation of polypeptide segments in proteins by systematic search. Proteins: Structure, Function and Genetics 1:46-163, 1986.

71. Gietzen, K., Sadorf, I., Bader, H. High affinity binding of the mastoparons by CaM. Biochem. J. 207:541-548, 1982.

72. Malencik, D.A., Anderson, S.R. A model for the regulation of the calmodulin-dependent enzymes erythrocyte Ca2+-transport ATPase and brain phosphodiesterase by activators and inhibitors. Biochem. Biophys. Res. Commun. 114:50-56, 1983.

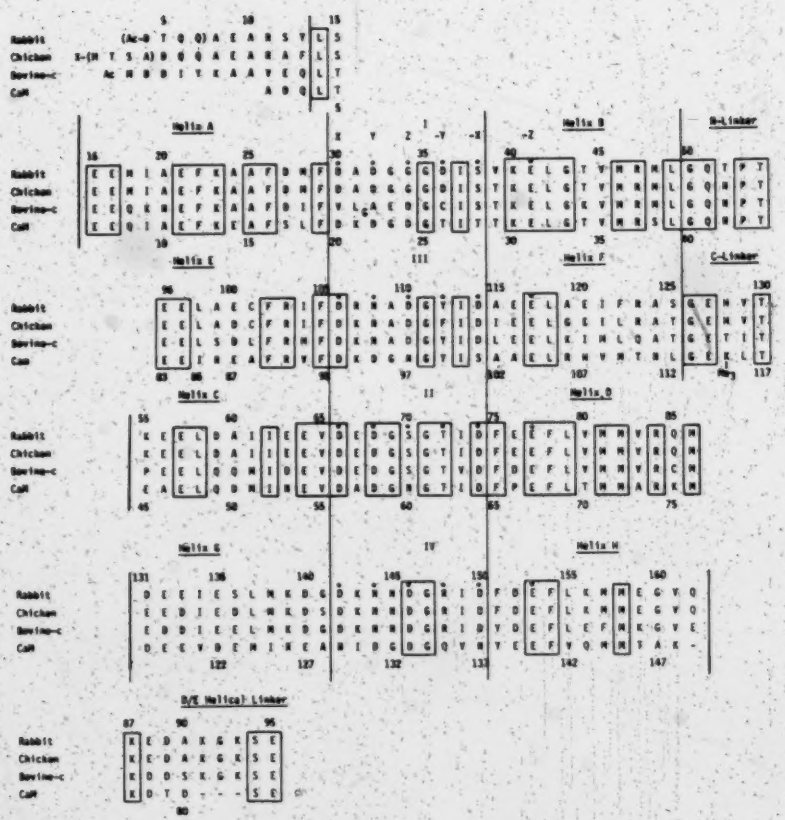


Fig. 4.1. The amino acid sequence alignment of calcium modulated proteins based on the tertiary structure of TnC. The sequence alignment of calcium modulated proteins based on the tertiary structure of TnC. The sequences are of rabbit skeletal TnC (64), chicken skeletal TnC (64,65), bovine cardiac TnC (65,66), and bovine brain CaM (67). The amino acid numbering is that of chicken skeletal TnC on top and bovine brain CaM below. The segments of the molecule are labelled according to the scheme given in Fig. 4.2a. Residues involved in coordinating to Ca^{2+} are indicated by an asterisk (*). The sequence comparison has been made so that the homologous pairs of EF hands are aligned, i.e., binding loop I with III and binding loop II with IV. The vertical lines demark the initiation and termination of helices. Note the ends of the loops and the following helices overlap by 3 residues. Those residues that are identical in all 4 proteins are enclosed in boxes.

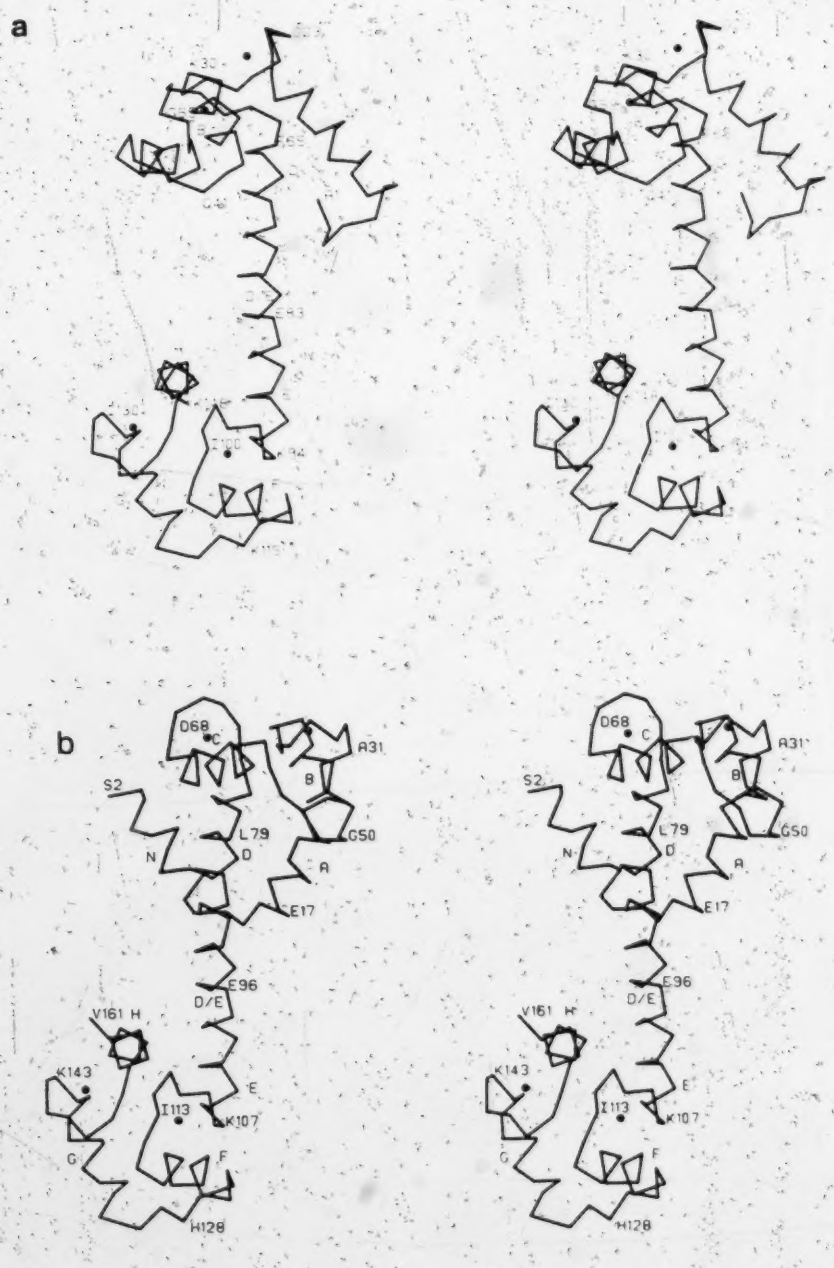


Fig. 4.2. (a) A stereo representation of the c-α backbone of the calcium-saturated ($4Ca^{2+}$) model of bovine brain CaM. Helices are labelled appropriately (A,B,C,D, D/E, E,F,G,H) and coincide with Fig. 4.1. Calciums are represented by black spheres. (b) A stereo view of the c-α backbone of calcium-saturated ($4Ca^{2+}$) turkey skeletal troponin C. Labelling as for Fig. 4.2a. For comparison, the orientation of the C-terminal domain is the same as that of $4Ca^{2+}$.

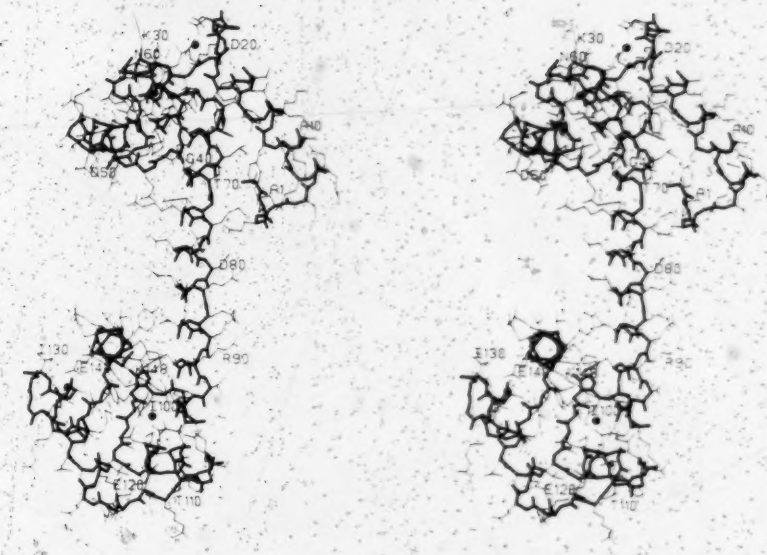


Fig. 4.3. An all atom stereographic stick representation of the $4Ca^{2+}$ CaM model. Main chain is in thick lines, side chains are in thin lines. Numbering is as in Fig. 4.1.

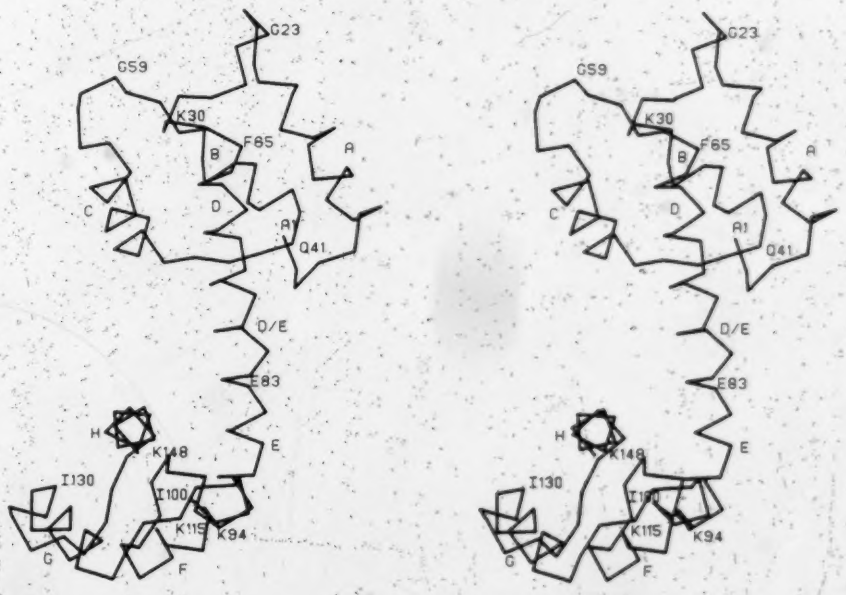


Fig. 4.4. A stereoview of the c-α backbone of the calcium-free ($0Ca^{2+}$) CaM model. Numbering and naming of helices is as in Fig. 4.1.

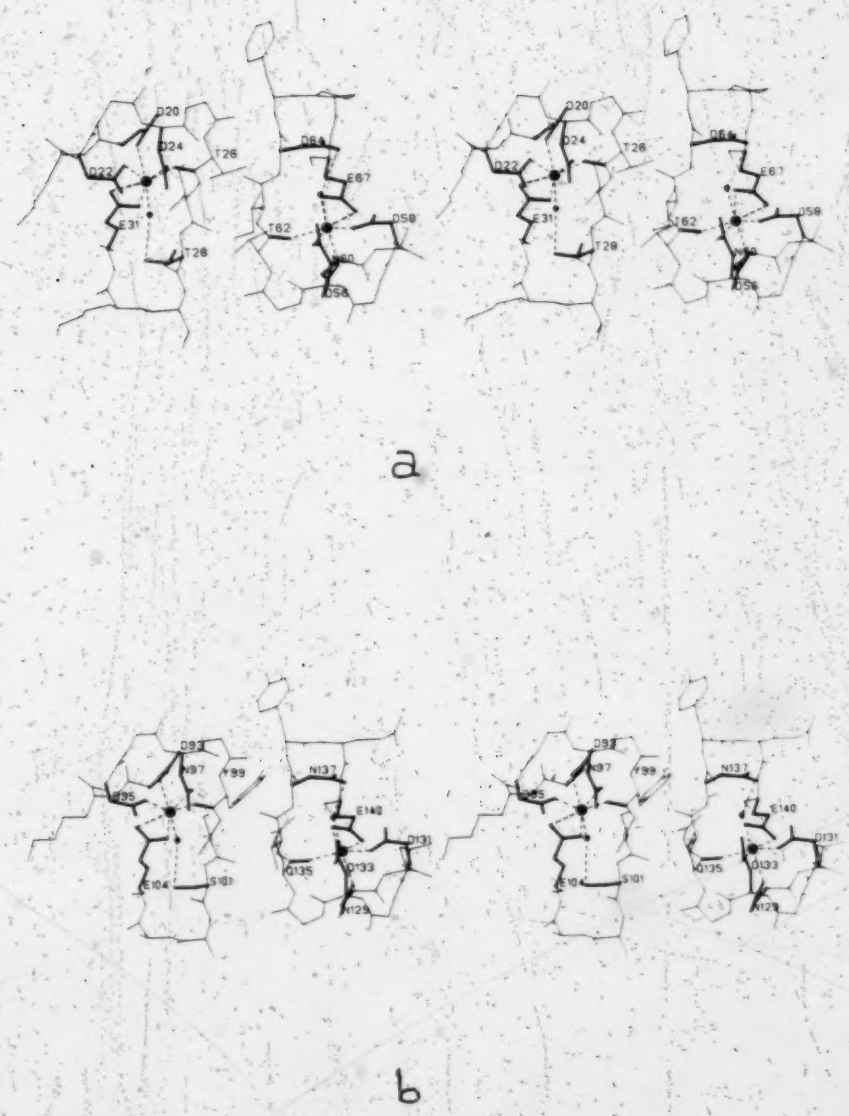


Fig. 4.5. (a) A stereo representation of calcium binding sites I and II in the $4Ca^{2+}$ CaM model. Calcium coordinating side chains or main chain carbonyls are shown thick. Calciums and waters are shown as large and small black spheres, respectively. Calcium-residue coordination is represented by a dashed line. (b) A stereo representation of calcium binding sites III and IV. Labelling is as for (a).

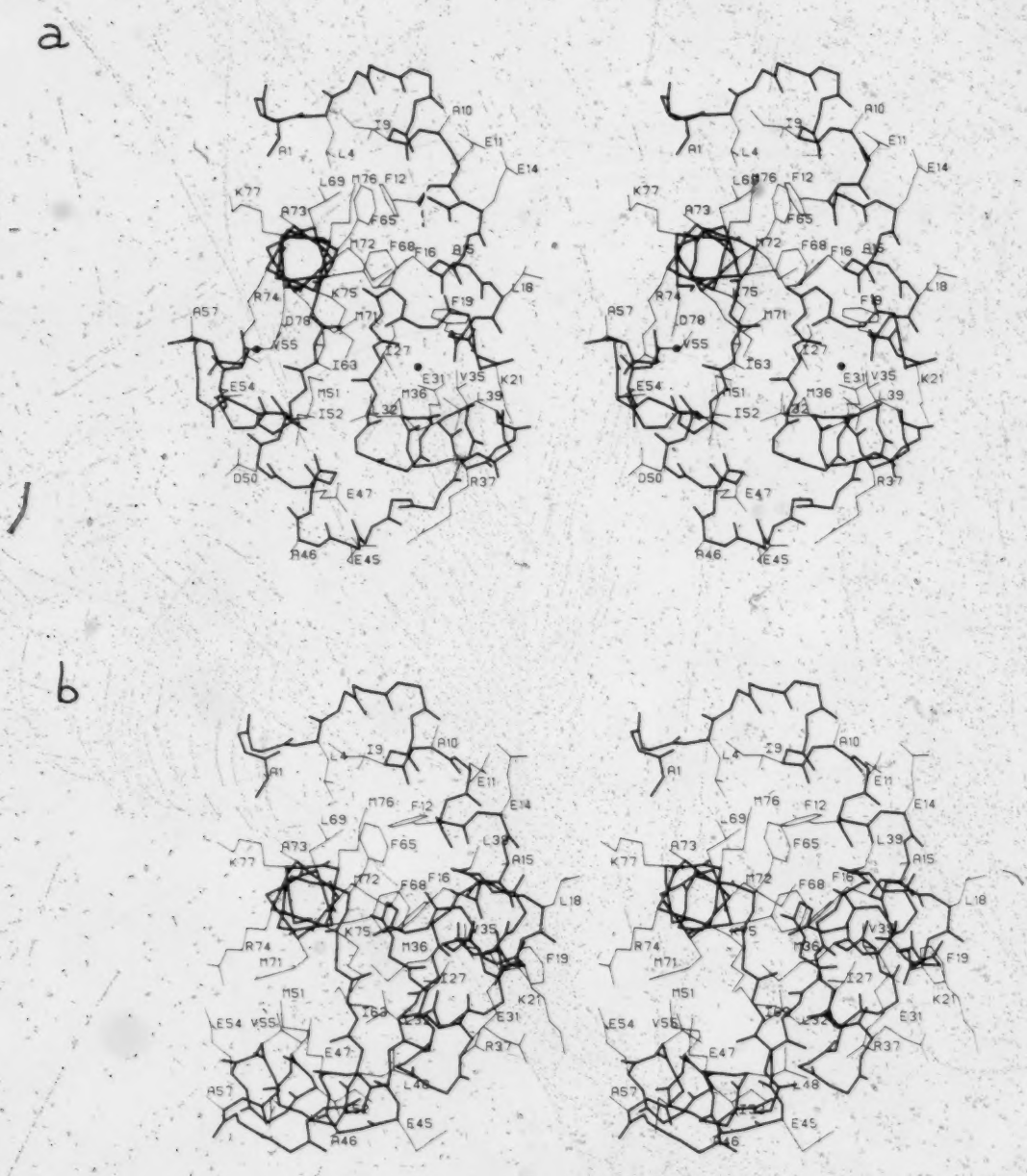


Fig. 4.6. (a) A stereo view of the N-terminal domain in the "open" $4Ca^{2+}$ CaM model. The orientation is looking down the D helix. Main chain atoms are thick lines, side chains are thin lines. Calcium ions are denoted as black spheres. For clarity, only hydrophobic residues and select charged residues are represented with side chains. (b) A stereo view of the N-terminal domain in the "closed" $0Ca^{2+}$ CaM model. The orientation is, again looking down the D helix.

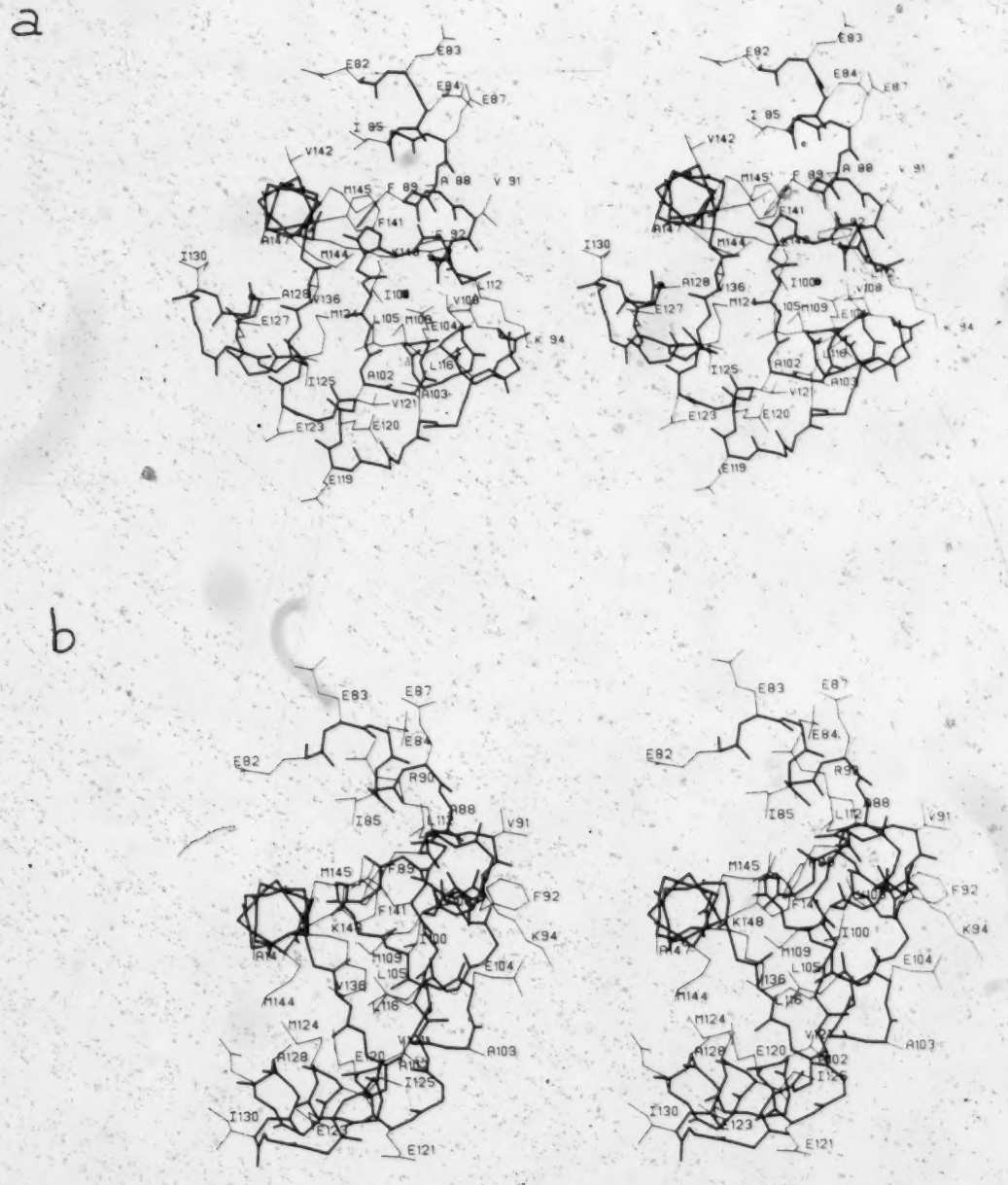
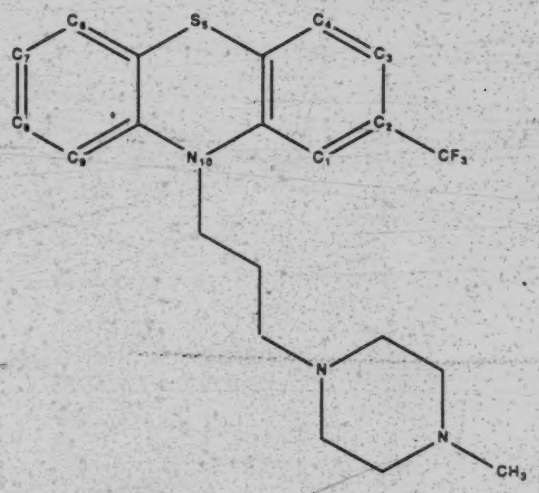


Fig. 4.7. (a) A stereo representation of the C-terminal domain in the "open" $4Ca^{2+}$ CaM model. The orientation is looking down the H helix, analogous to the view in Fig. 4.6a. Main chain atoms are thick lines, side chains are thin lines. Calcium ions are denoted as black spheres. For clarity, only side chains for hydrophobic and select charged residues have been represented. (b) A stereo view of the C-terminal domain in the "closed" $0Ca^{2+}$ CaM model. Orientation and labelling is as for (a).

a



b

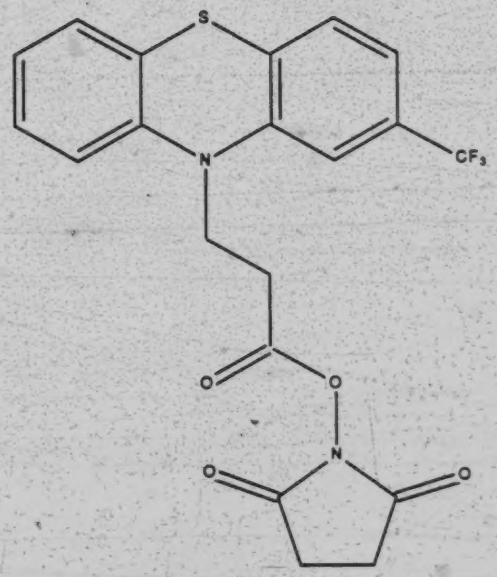
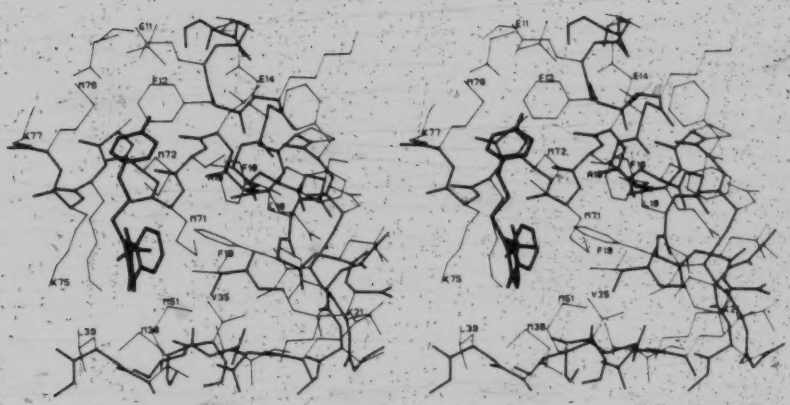


Fig. 4.8. (a) The chemical configuration of trifluoperazine (TFP). Molecular coordinates used for modelling were those of McDowell (68). (b) The chemical configuration of 10-(3-propionyloxysuccinimide)-2-(trifluoromethyl)phenothiazine (POS-TP) (69).

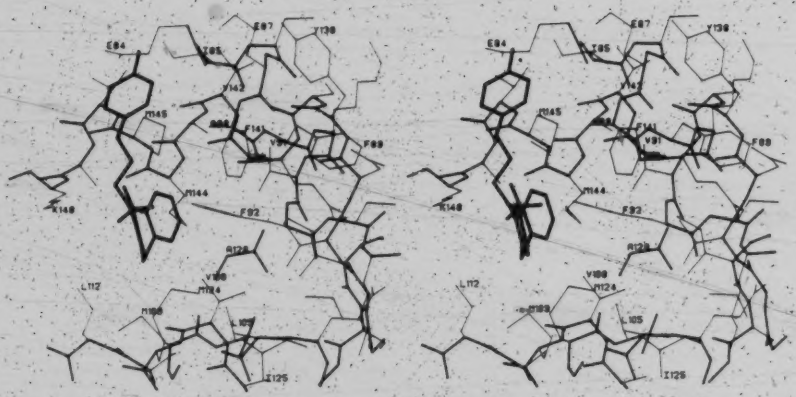
a



b

Fig. 4.9. (a) A stereographic stick model of the binding of TFP to the N terminal domain of the $4Ca^{2+}$ CaM model. Main chain and the TFP molecule are in thick lines, side chains are in thin lines. (b) A space filling picture of TFP bound in the N terminal domain of $4Ca^{2+}$ CaM. Amino acid side chains are in standard shapely colors. Negatively charged groups in reds, positively charged groups in blues, methionines yellow, hydrophobes are green, aromatics are brown. The view is approximately 90° to that in (a).

a



b

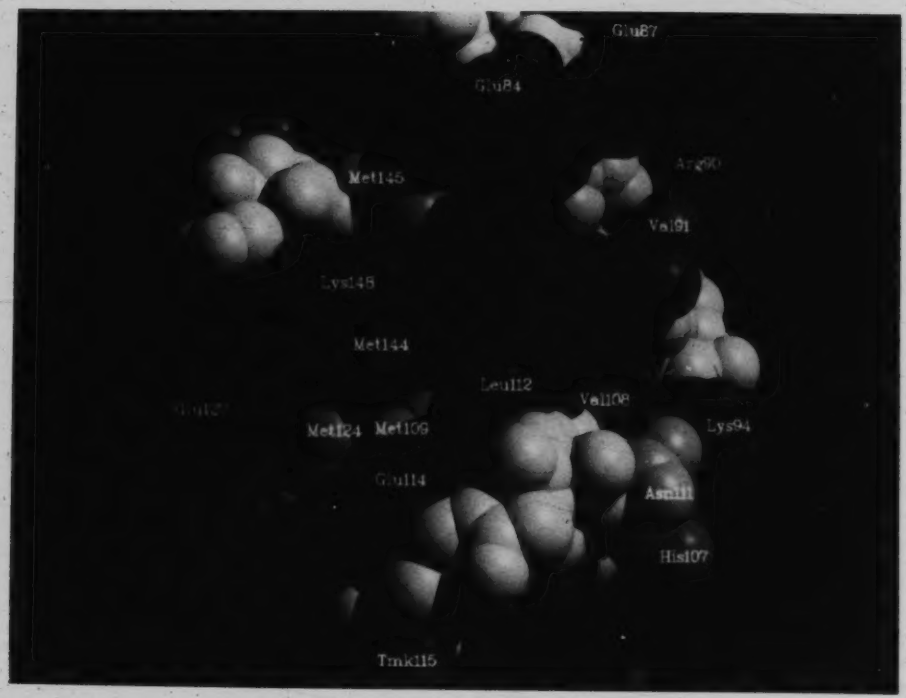


Fig. 4.10. (a) A stereographic stick model of the binding of TFP to the C terminal domain of the $4Ca^{2+}$ CaM model. Main chain and the TFP molecule are in thick lines, side chains are in thin lines. (b) A space filling picture of TFP bound in the C-terminal domain. Amino acid colors as for Fig. 4.9b. The view is approximately 90° to that in (a).

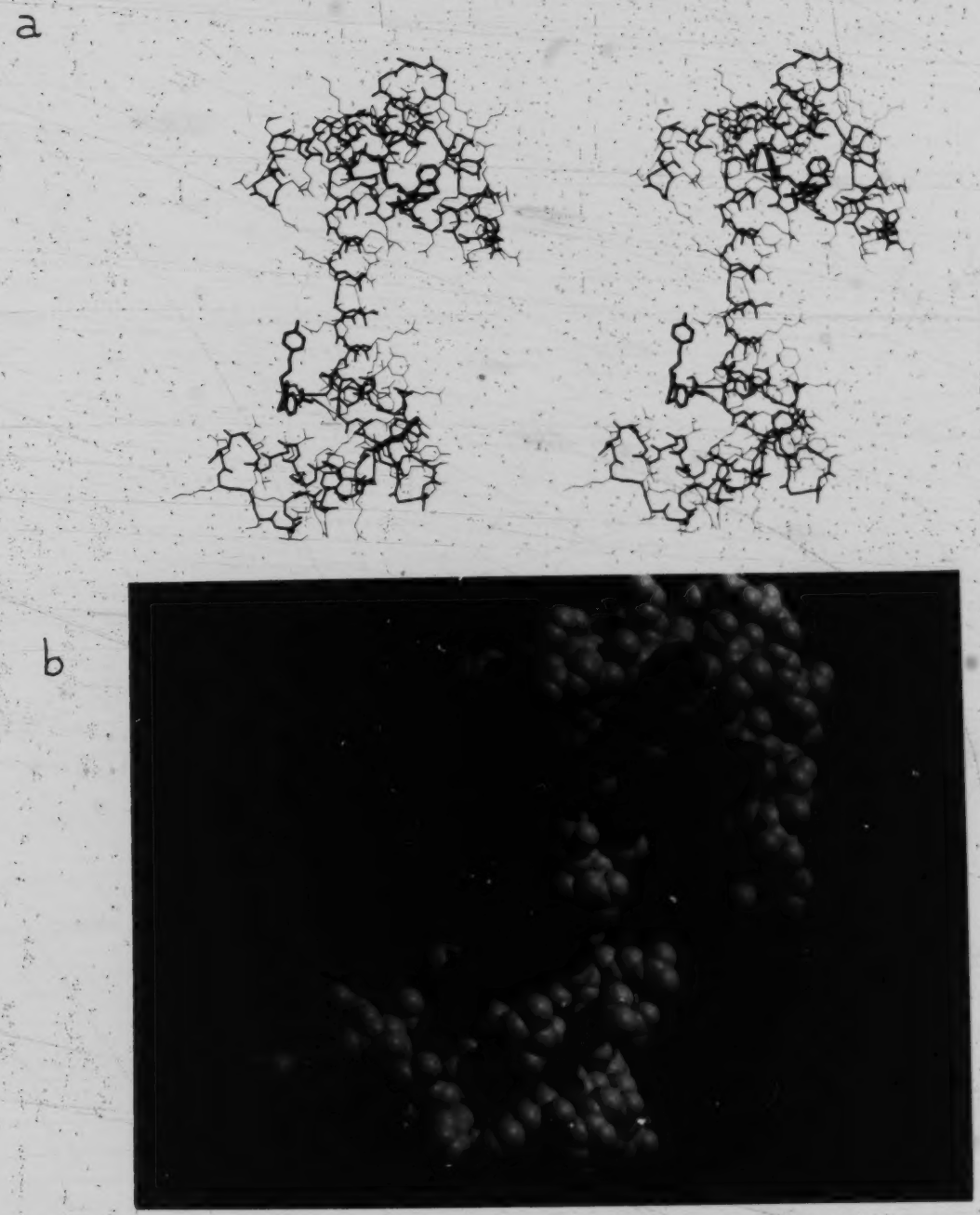


Fig. 4.11. (a) A stereo view of the entire calcium-saturated model showing the two proposed binding sites for the antipsychotic drug trifluoperazine. Main chain and TFP molecules are in thick lines, side chains are in thin lines. (b) A view of calmodulin sharing the two trifluoperazine binding sites. The calmodulin atoms are represented as white spheres with the glutamic acid residues involved in binding the piperazine moiety of TFP in red. The TFP molecules are shown as blue spheres. This figure as well as Figs. 4.9b and 4.10b were made using "raster 3 D" written by David Bacon.

CHAPTER 5

A MODEL FOR THE INTERACTION OF AMPHIPHILIC HELICES WITH TROPONIN C AND CALMODULIN[1]

INTRODUCTION

Troponin-C (TnC) and calmodulin (CaM) belong to the helix-loop-helix (HLH) Ca^{2+}-binding protein family (1). Each protein has four analogous HLH structural motifs that bind Ca^{2+} ions with K_ds ranging from 10^{-5} to 10^{-7} M (2-5). The binding of calcium to the loops of the HLH units induces the conformational changes in TnC and CaM that are required for the optimal interactions with their target molecules. The effect of the conformational change in TnC is transmitted to the thin filament proteins troponin I, tropomyosin and actin thereby triggering muscle contraction (For recent reviews of the biochemical data, see refs. 2,3). CaM binds to and modulates the activities of a variety of diverse intracellular enzymes including myosin light chain kinase, calcineurin, erythrocyte Ca^{2+}-ATPase, brain adenylate cyclase and cyclic 3',5' monophosphate phosphodiesterase (4,5).

The high resolution crystal structures of turkey TnC (6), chicken TnC (7), and bovine brain CaM (8), have been determined. These crystal structures revealed dumbbell-shaped molecules having 2 globular domains (the N and C-terminal domains) joined by a long helical linker (Fig. 5.1). Each globular domain is cup-shaped. From this analogy, the bottom of the cup contains the pair of Ca^{2+} binding loops, the outside and upper rim of the cup is comprised mainly of negatively charged residues, and the interior of the cup is lined with hydrophobic residues (Fig. 5.2a and 5.2b). In the C-terminal domain of TnC and in both the N and C-terminal domains of CaM, the Ca^{2+}-binding sites are occupied by metal ions; the resulting conformations of these domains are similiar (9). In them, the hydrophobic residues that line the interior of the cup are relatively exposed (Fig. 5.3a, b,

---

[1] A version of this chapter has been published. Strynadka & James 1990. *Proteins: Structure, Function and Genetics* 7:234-248. Relevant material published subsequent to the paper is given in Chapter 7.

c). However, due to the low pH of the crystallization conditions (7,10), the N-terminal domain of TnC has no ions in the expected coordination sites and consequently adopts a very different tertiary arrangement (11). Thus, in the absence of bound Ca^{2+}, the N-terminal domain is much more "closed" with the interior hydrophobic residues buried away from the protein/solvent interface (Fig. 5.3d).

Because of the similiar conformations of the Ca^{2+}-filled domains of TnC and CaM and of the extensive analogy between the N and C-terminal domains of TnC, it seemed probable that the N-terminal domain of TnC should adopt a fold similiar to that of the Ca^{2+}-filled state. A model for the interconversion of the Ca^{2+}-free (closed) and Ca^{2+}-bound (open) forms of the N-terminal domain of TnC has been proposed (11). The model is supported by a number of solution studies that implicate the movement of several hydrophobic residues to a more solvent exposed state upon calcium binding (12-16).

The calcium-bound conformations of TnC and CaM are required for high affinity binding to target molecules (2-5). Several studies indicate that the hydrophobic surfaces of TnC and CaM that become exposed upon calcium-binding form the sites of target molecule interaction (13-26). Although the details of these interactions have not been defined at the molecular level, it is clear that both hydrophobic and electrostatic interactions are involved. Additionally, there is increasing evidence that amphiphilic helices may play a major role in target molecule binding. The proposed CaM binding sites on myosin light chain kinase (27,28), phosphofructokinase (29), phosphorylase b kinase (28,30) and calmodulin-dependent protein kinase from brain (31), all encompass short sections of each enzyme that have high amphiphilic helix forming propensities. As well, many known potent inhibitors of CaM and TnC are short amphiphilic helical peptides (for a recent review see 32). Ostensibly, their inhibitory action is through competitive binding with target molecules for the exposed hydrophobic regions on TnC and CaM.

Analysis of the crystal packing of turkey skeletal TnC provides a possible model for the binding of amphiphilic helices to the hydrophobic clefts of TnC and CaM. The

molecular packing of the native turkey skeletal TnC in the trigonal crystals shows that the A helix of one TnC molecule fits into the open Ca^{2+}-bound C-terminal hydrophobic cleft of a symmetry related molecule (Fig. 5.4). We have used this helix as a template for our modelling. The following describes our analysis of the modelled TnC and CaM complexes with mastoporan and melittin.

METHODS

The A helix (for helix nomenclature see ref. 9) of the molecule at -x+y, 1-x, z+1/3 in the crystal structure of native turkey skeletal troponin C was used as a template for the modeling of amphiphilic peptides into the Ca^{2+}-bound hydrophobic cleft on the adjacent symmetry related neighbor at x,y,z. These molecules are denoted in Fig. 5.4 as molecules 2 and 1, respectively. The program MUTATE (R. Read, unpublished) was used to substitute the appropriate residues of mastoporan (33) and melittin (34) into the TnC A-helix coordinate set. The sequence alignment used for these substitutions is given in Table 5.1. Molecular surfaces to show van der Waal contacts were calculated using the algorithm of Lee and Richards (35). Side chain clashes that violated van der Waals contact distances were minimized using the mms suite of programs (S. Dempsey) on a Silicon Graphics Iris 3030 workstation.

The last three residues of the H helix in TnC are distorted from a helical conformation in the native structure due to crystal packing constraints (see Fig. 5.4 and Fig. 5.5a). In order to deduce possible interactions that the H-helix of TnC might have with the modelled amphiphilic peptides, Glu159-Val161 were given a regular α-helical conformation, so that the H helix is continuous along its entire length.

For the modelling of the mastoporan and melittin interactions with bovine brain CaM, least-squares overlap methods were used to transform the A-helix coordinate set described above into both calcium saturated domains of the bovine brain calmodulin crystal structure (8). Modelling then proceeded as detailed for TnC.

RESULTS AND DISCUSSION

Analysis of the Crystal Packing of Troponin C

Turkey skeletal TnC crystallizes in the trigonal space group $P3_221$ with unit cell dimensions $a,b = 50.54$ Å, $c = 66.91$ Å, $\gamma = 120°$ (6). There are six TnC molecules within the unit cell. The packing interaction of interest here involves the contacts between the molecule at x,y,z and the symmetry related one at -x+y, 1-x, z+1/3. Figs. 5.4 and 5.5a and the data in Table 5.2 clearly show that the A helix in the N-terminal domain of the molecule at -x+y, 1-x, z+1/3 (molecule 2 in Fig. 5.4) fits neatly into the C-terminal hydrophobic pocket of the molecule at x,y,z. The C-terminal hydrophobic patch, formed by residues on helices E, F, the F-G linker, and helices G and H is in the Ca^{2+}-filled or "open" conformation in the native structure i.e. its hydrophobic residues are relatively exposed to solvent (Fig. 5.3a). In general, large areas of solvent-exposed hydrophobic side chains are considered energetically unfavourable. Perhaps the packing of the relatively uncharged face of the A-helix into the exposed hydrophobic pocket serves to minimize this unfavorable energy in the crystallization process. It is also tempting to speculate that this interaction may be responsible for the pH dependent dimerization of TnC molecules in solution. Aggregation of TnC occurs only at acidic pH's that also prohibits metal binding to the low affinity N-terminal Ca^{2+}-binding sites (36). This would suggest that the 2 Ca^{2+} form is required for dimerization to occur. Furthermore, the three-dimensional structures of turkey and chicken TnC show that under the low pH of the crystallization conditions (~4.8), the protonation of the carboxylate groups of acidic side chains allows for several strong carboxyl/ carboxylate interactions which not only stabilize the Ca^{2+} free form of the N-terminus within each molecule, but also stabilize the interaction of the two symmetry related molecules shown in Fig. 5.4.

Analysis of the X-ray Crystallographically Defined Interaction of the A-Helix with the C-Terminal Domain of Troponin C

Hydrophobic interactions that stabilize the A-helix within the environment of the

neighboring C-terminal hydrophobic cleft can be summarized as follows: The A-helix presents a relatively hydrophobic face to the adjacent apolar pocket (Ala20, Ala24, Ala25, Met28; see Fig. 5.5a and Table 5.1). Because of the small surface area of alanine side chains, most of these nonpolar interactions are probably weak. The Cβ of Ala20 faces in towards the hydrophobic pocket where it is close to Leu122, Val129, and the Cβ of Glu127. Ala24 Cβ makes contact with the side chains of Phe105, Ile121 and Leu122. The Ala24 contact distances are quite long and observation of the calculated molecular surfaces shows the presence of a cavity in this region. By far the most extensive intermolecular hydrophobic interactions arise from Met28 of the A-helix. Met28 fits snugly into a pocket formed by Phe105 and Ile104. Met28 is also stabilized by interactions with Phe29 from the A-helix itself. All other hydrophobic residues of the A-helix (Table 5.1) lie on the opposite face where they form intramolecular hydrophobic interactions within the N-terminal hydrophobic domain of their own molecule.

The packing of the A-helix into the neighboring C-terminal domain is also enhanced by several electrostatic interactions which occur along the rim of the hydrophobic cup. Glu16 on the A-helix forms a carboxyl-carboxylate interaction with the side chain of Glu127 on the F-G linker (Glu16 OE1-Glu127 OE2, 2.3 Å) of the symmetry related molecule. In addition, a network of water molecules interacts with both Glu16 and Glu127 (Glu16 OE1-Wat197, 3.3 Å; Glu127 OE1-Wat205, 2.8 Å). As well, Glu17 forms a carboxyl-carboxylate interaction with Glu133* on the G-helix (Glu17 OE1-Glu133 OE1, 3.3 Å).

On the whole, the A-helix provides a weak hydrophobic face to the adjacent apolar cleft and the binding is also stabilized by electrostatic interactions on the rim of the cleft. A glaring exception to note is the side chain of Glu21 on the A-helix which points directly at the apolar pocket. To offset this potentially unfavorable interaction, an elaborate system of

*The final electron density map of turkey TnC is more consistent with a glutamate at position 133 rather than aspartate as in chicken. This has been confirmed by amino acid sequence studies (M.R. Carpenter, C.K. Golosinki & L.B. Smillie, unpublished data).

water molecules surrounding the negatively charged side chain probably serves to keep it relatively happy in its hydrophobic surroundings. Glu21 OE2 interacts with Wat261 (3.2 Å), in turn Wat261 forms a hydrogen bond with Wat227 (2.2 Å), and Wat227 contacts Wat199 (3.3 Å) (see Fig. 5.5a).

Analysis of the Modelled Interaction of Mastoporan with the C-Terminal Domain of Troponin C

TnC and CaM bind a variety of naturally occurring extracellular peptides including the opiates β-endorphin (37) and dynorphin (38), various members of the glucagon family (39,40) and a group of cytotoxic peptides, among them bee venom melittin (34,41,42) and the mastoporans [(38,43), Table 5.1]. Binding of these peptides to calcium saturated TnC and CaM is typically with high affinity ($K_d = 10^{-8}$-10^{-9} M) and has been shown to inhibit effectively the modulating activity of these two HLH proteins for their intracellular targets (40,41,43-47,49).

There are two characteristics common to all the above TnC and CaM inhibitory peptides. Firstly, all contain a sequence with high α-helix forming propensity [(48), Table 5.1]. As well, the hydrophobic and hydrophilic residues are arranged in the sequence such that hydrophobic residues would extend off one face of the helix, and charged, hydrophilic residues from the opposite face; i.e., they are amphiphilic in nature. Solution circular dichroism and NMR studies involving β-endorphin (40), mastoporan (40,43,50,51) and melittin (42,44,52) indicated an induction of helical conformations in these peptides as they formed complexes with the complementary surface of CaM. Direct evidence for the amphiphilic helical nature of melittin comes from the X-ray crystallographic structure (53). Finally, DeGrado and coworkers have designed from first principles short amphiphilic helices which bind CaM with submicromolar dissociation constants (32,48). One particular peptide was synthesized with an environmentally-sensitive fluorescent amino acid (Trp) sequentially placed at each position along the chain. The pattern of periodic fluorescence decay of this set of analogues in complex with CaM could be explained by an

α-helical structure for the bound peptide (54).

If, as all the above evidence suggests, the peptides in Table 5.1 are amphiphilic helices then it is clear that they would form a significant hydrophobic area, and thus, in turn would require a large complementary hydrophobic area on TnC or CaM with which to interact favorably. From the crystallographic structures (6-8) it appears that only the exposed hydrophobic pockets found in the calcium-bound domains could accommodate these helices.

The second property common to all the TnC/CaM inhibitory peptides is the presence of a cluster of residues that are positively charged at neutral pH (Table 5.1). The importance of electrostatic charge interactions in peptide binding has been implicated in several previous studies. For example, the affinity of β-endorphin for CaM is significantly decreased as ionic strength is increased (39). As well, amphiphilic peptide inhibitors designed with acidic instead of basic residues comprising their hydrophilic face failed to bind to CaM (48).

Mastoporan, mastoporan X, and mastoporan P constitute a family of naturally occurring tetradecapeptides isolated from the vespid wasp (33). As mastoporan is considered to be one of the most potent peptide inhibitors of both TnC and CaM it seemed a logical first choice for our modelling study.

The appropriate residues of mastoporan were aligned with the A-helix of TnC as shown in Table 5.1. The modelled mastoporan provides a hydrophobic face to pack into the C-terminal hydrophobic cleft of TnC, with potential electrostatic interactions occurring along the outsides of the pocket. However, whereas the hydrophobic interactions in the A-helix/TnC case seem weak because they are from alanyl side chains predominantly, the potential hydrophobic interactions in the binding of mastoporan are more extensive and undoubtedly play a major role in the high affinity binding of this molecule for TnC. Figs. 5.2c and 5.5b depict the structural fit of the mastoporan helix into the hydrophobic groove on the C-terminal domain of TnC and Table 5.2 gives a summary of the proposed

potential hydrophobic and electrostatic interactions.

Mastoporan provides a total of seven hydrophobic residues that could potentially interact with the hydrophobic residues in the cleft of the C-terminal domain. The positively charged amino terminus of the modelled drug is favourably close to the negatively charged cluster of residues, E133 and D136. Ile1 contacts Val129 as well as the aliphatic portions of the side chains of Glu127 and Glu133. Asn 2 is near Glu133 and Asp136, both of which protrude from the lip of the hydrophobic pocket. The C^{β} of Asn2 is near the C^{β} of Leu137. Leu3 forms a hydrophobic network with Leu137 and Val161 that extends from the H-helix of TnC. By conferring a continuous α-helical conformation upon helix H in TnC, Val 161 now provides potential extensive hydrophobic interactions with residues of mastoporan (Leu3, Leu6). It is interesting to note that a valine at position 161 is highly conserved among all the sequenced troponin C's (55-57); the equivalent residue in CaM is Lys148. Lys4 extends to the outside of the mastoporan amphiphilic helix where it is in a favourable position to form an electrostatic interaction with the side chain of Glu127 that extends from the F-G linker in TnC. The proposed alignment shown in Table 5.1 indicates a lysine at this position in the many known amphiphilic peptides that bind TnC and CaM as well as the highly conserved presence of a glutamic acid residue at position 127 in both TnC and CaM. Ala5 of mastoporan interacts with Val129 on the F-G linker peptide. Leu6 is a key residue in the mastoporan/TnC interaction as it forms favorable interactions with Met157, Met158, Leu137, and Val161 (Table 5.2). Table 5.1 indicates that in the alignment proposed, this position is invariably a hydrophobic residue, often a Leu or Ile in other amphiphilic inhibitory peptides. Ala7 and 8 both point to the outside of the helix. Ala7 is relatively exposed to solvent, whereas Ala8 is more protected with its Cβ pointing in at the Cβ of Thr125. Leu9 is another key contact residue with close hydrophobic contacts to Phe105 and Ile121. Ala10 interacts weakly with Met158 (4.0 Å). Mutation to a slightly larger side chain at this position might increase binding by filling in the small cavity around this alanine. Lys11 and Lys12 both extend into the solvent on the hydrophilic side

of the mastoporan helix. The positively charged ammonium ion of lysine 12 effectively "caps" the F helix at its C-terminus, the most negative portion of its helix dipole. Ile13 binds to Ile104 and Phe105. Leu14, in turn, interacts favorably with the adjacent Ile13 in the mastoporan itself.

Analysis of the Modelled Interaction of Mastoporan with the C-Terminal Domain of Calmodulin

The extensive sequence and structural analogy of TnC and CaM is well documented and was the basis of our earlier computer modelling of CaM based on the structure of turkey skeletal TnC (19). Superpositioning of the C^{α} atoms of the C-terminii of TnC and CaM yields an r.m.s. of less than 1.0 Å (9). This fact, in combination with the high degree of conservation of hydrophobic residues within the hydrophobic pockets of TnC and CaM (compare Fig. 5.3a and 5.3c) would suggest that mastoporan could bind in similar positions in the two molecules. Residues involved in our model of mastoporan binding to CaM (Fig. 5.5d) (and the analogous residue in TnC in brackets) include the following: Ala88 (Cys101), Val91 (Ile104), Phe92 (Phe105), Val108 (Ile121), Met109 (Leu122), Leu112 (Thr125), Glu114 (Glu127), Glu123 (Asp136), Met124 (Leu137), Met144 (Met157) and Met145 (Met158). In addition the alkyl side chain of Lys148 (Val161) could interact with Leu3 of mastoporan. The replacement of an Ala for a Cys at position 88, of a Met for a Leu at position 109, and of a Leu for a Thr at position 112 serves to increase the hydrophobic surface area in CaM with which mastoporan may interact.

The smaller interhelical angles of the HLH unit in calmodulin [as compared to TnC (9)] serves to promote closer fit of the amphiphilic helices into the hydrophobic patch. This is mainly due to the altered position of the loop joining the two helices (compare Fig. 5.5b and 5.5d). This phenomenon, in addition to the increased hydrophobic interaction area mentioned above, may contribute to calmodulin's greater association with mastoporan in comparison to TnC.

A differential trace labelling study of lysine residues in CaM showed that binding of

either an amphiphilic inhibitory peptide (β-endorphin) or the antisphycotic drug TFP decreased the reactivities of Lys75 and Lys148 most dramatically (58), thus suggesting common binding sites for the two molecules. Lys75 extends across the N-terminal hydrophobic domain; Lys148, although not seen in the crystal structure (8) is thought to lie in an analogous manner near the C-terminal hydrophobic domain (19). Indeed, several of the residues in CaM proposed to be involved in binding mastoporan are also those proposed to be involved in the binding of TFP to CaM (19). This observation would help explain the competitive effect of TFP on the binding of these amphiphilic helices to CaM (13,19,38,39). Furthermore, as discussed in the earlier TFP modelling study, several of these hydrophobic residues on CaM as well as Lys75 and Lys148 would be substantially buried in the Ca^{2+}-filled (open) to Ca^{2+}-free (closed) transition (19). Therefore, as with the case of TFP binding, one would also expect a decreased affinity of the modelled amphiphilic helical mastoporan peptide for CaM as the accessible hydrophobic surface with which it could interact is severely diminished in the Ca^{2+}-free state. This would explain the calcium requirement for the high affinity interaction of the amphiphilic helices to CaM and TnC.

Binding of Amphiphilic Helices to the N-Terminal Domain of CaM

Although different experimental conditions have led to conflicting reports there is evidence that 2 moles of mastoporan (59,60), are bound per mole CaM; one with high affinity and one with lower affinity. A ^{113}Cd NMR study found that two moles of mastoporan/mole of CaM broadened the four ^{113}Cd NMR signals over that found in a 1:1 molar ratio without affecting the chemical shifts (59). This observation was interpreted as binding of a second molecule of mastoporan with lower affinity than the first; the absence of changes in the ^{113}Cd NMR chemical shifts argues against major conformational changes in calmodulin (at least near the metal binding sites), as the second peptide molecule binds to the protein. Proteolytic digestion studies of CaM showed that the C-terminal 78-148 fragment bound mastoporan with an affinity equal to that of the intact protein (60).

The N-terminal 1-77 fragment also bound mastoporan, but with reduced affinity (K_a 10^5 M^{-1}). This reduced affinity may be due to disruption of the D-helix. Indeed, it had been shown earlier that the 1-106 fragment of CaM bound mastoporan with higher affinity (38). This example highlights the inherent limitations of tryptic digestion studies in defining molecular binding sites. The choice of fragmentation is often limited, and may give negative results simply because the tertiary structure of the binding area is being destabilized by the removal of adjacent secondary structural units. For instance the 107-148 tryptic fragment of CaM was found to bind very weakly to mastoporan and the area was subsequently dismissed as not being the major site for drug binding (38). However, perusal of the CaM structure shows that removal of amino acids 85-106 [which includes all of Ca^{2+}-binding loop III and several key residues of the domains hydrophobic core (9)] would be devastating to the conformation of the C-terminal domain.

That mastoporan would bind simultaneously to both the N and C-terminal domains of CaM might be expected considering the extensive homology between the 2 domains, especially in the hydrophobic pocket regions where we propose that the peptides will bind (compare Fig. 5.3b and 5.3c). Mastoporan is small enough that binding to one domain would not interfere with binding at the second. All of the hydrophobic residues of CaM involved in peptide complex formation in the C-terminal domain have analogous counterparts in the N-terminal domain. A notable difference in the N-terminal binding site is a smaller number of negatively charged side chains surrounding the hydrophobic pocket. Most significantly, there is no counterpart to Glu114 of the C-terminal domain which is in a favourable position to interact with the lysine extending off the mastoporan helix at position 4. This in part may account for the lower affinity of peptide for the N-terminal domain of CaM (60).

Recently, computer molecular-modelling study involving calmodulin and its interactions with the calmodulin-binding domain peptide from MLCK has been described (61). In that model, the 27 amino-acid peptide was given an α-helical conformation and the

authors proposed that it interacts simultaneously with both the N-terminal and C-terminal hydrophobic patches of CaM. This work was a logical extension of the earlier crosslinking study (62) that suggested the central helix of CaM could act as a "flexible tether" to accommodate interactions with a variety of target enzymes. Subsequently, support for this modelling has come from a series of deletion mutants of CaM that retain their activation ability for some targets in spite of the removal of segments of the central helix (63).

In order to accomplish the MLCK peptide-CaM complex model building, the authors introduced a non-helical bend (ϕ = -54, ψ = +98) into the central linker helix of CaM at Ser81. This single change was not sufficient to achieve the required interactions so smaller adjustments were made to the main-chain torsional angles of several residues adjacent to Ser81. The resulting proposed conformational change brings the hydrophobic patch on each domain to face towards one another so that hydrophobic residues on one side of the MLCK peptide would interact with the N-terminal domain and at the same time other hydrophobic residues on the opposite side of the peptide would interact with the C-terminal domain.

In its present form, we found that the bent helix model (61) could not account for simultaneous interactions of residues on a single helical peptide with the hydrophobic regions of both domains of CaM. The average distance between hydrophobic residues on the opposing N- and C-terminal patches is approximately 20 Å, a distance far too large to allow for effective hydrophobic interactions (~ 3.8-4.2 Å) from both domains to the same helical peptide. In addition to this disparity in the interdomain distances there are several other reasons why such a model can not easily reconcile the high affinity binding of mastoporan to CaM.

Firstly, as mentioned above, the isolated C-terminal tryptic peptide of CaM has an affinity for mastoporan that is almost equal to that of intact CaM (60). The bent-helix model would predict a quite different affinity between these two forms of CaM which is clearly not the case. Secondly, even if it were possible to bring the two domains

sufficiently close together so they could interact simultaneously with a single peptide, the strongly amphiphilic character of mastoporan would require one of the domains to interact with a predominantly polar/charged surface of the helical peptide. Certainly charged residues such as lysines can contribute to hydrophobic interactions via the aliphatic portion of their side chains, but their ionic moieties must still somehow be satisfied in the apolar environment. Table 5.1 indicates that the charged face of mastoporan would present three such moieties is (Lys4, Lys11, Lys12) to one of the hydrophobic domains of CaM in the bent helix model. In light of the similarity of the N and C-terminal hydrophobic domains of CaM, and the small size of mastoporan, it is difficult to reconcile why it would be more energetically favourable to distort the central helix of CaM so that one of the hydrophobic domains would interact with a charged face of a single peptide rather than to bind the hydrophobic face of a second peptide molecule.

A Model for the Interaction of Melittin Binding to CaM

Like mastoporan, bee venom melittin complexes with high affinity (10^9 M^{-1}) to CaM and contains a 14 residue segment with an amphiphilic, helical nature (34,41,42,44, 52,53,64,65). However, melittin differs from mastoporan in that this 14 residue segment is embedded within a larger 26 residue peptide (Table 5.1). Additionally melittin binds CaM with 1:1 stoichiometry. Binding of one mole of melittin to one mole of CaM affects residues in both the N and C-terminal domains (52). Finally, melittin can complex with both the 1-77 and the 78-148 tryptic fragments but with reduced affinity for both (.5 μM and 2μM, respectively) (66).

The sequence alignment of a 14 residue amphiphilic segment of melittin to the A helix of TnC and the mastoporans is given in Table 5.1. In this alignment, the key TnC/CaM binding residues of mastoporan, Leu3, Lys4, Leu6, Leu9 are also retained in melittin (Leu6, Lys7, Leu9, and Leu13). Indeed we were able to construct a preliminary model of this 14 residue section of melittin complexed with TnC and CaM which is strikingly similar to that of mastoporan (see Table 5.2, Fig. 5.5c). As for mastoporan we would predict that

the hydrophobic face of this 14 residue segment of melittin could interact with analogous hydrophobic residues in either the N or C-terminal domains of CaM.

Although the binding of this 14 residue portion of melittin seems reasonable, especially in light of its strong homology with the mastoporan/CaM case, our model at this point did not account for the additional C-terminal 10 residues on melittin.

Judging from the cluster of positively charged basic residues in the C-terminus of melittin and the proximity of a cluster of negatively charged acidic residues on the D/E linker of CaM and TnC it seemed that this would be a favourable interaction site. Fluorescence labelling and tryptic digestion studies have also implicated the D/E linker region as important in binding melittin (38,64,66). If the position of the amphiphilic 14-residue segment of melittin in the hydrophobic cleft is to be maintained, then it is obvious that a bend in the melittin molecule would be required to redirect the positively charged C-terminal portion up along the D/E helical linker. Flexibility of the main chain dihedrals of Gly12 and Pro14 in melittin could favor such a bend. Indeed, a kinked conformation in the melittin helix was seen in the X-ray crystallographic structure solution of the isolated molecule (53).

With this in mind, we then proceeded to model the full melittin molecule onto CaM. What was immediately evident however was that in order to maintain hydrophobic interactions of melittin with the C-terminal apolar cleft and to accommodate interactions of the C-terminal tail with the D/E helix of CaM required a stereochemically unreasonable bend at Gly12/Pro14 of melittin. We found that the only way to position a more reasonably curved melittin helix was to rotate the N-terminal segment of melittin by 180° around its helix axis. This movement allows the dipole of the C-terminal helix of melittin to run in a favourable manner relative to the D/E helix of CaM, and allows contact of oppositely charged residues on the two molecules. However, this rotation also serves to place the partly hydrophilic face of the N-terminal amphiphilic segment of melittin into the C-terminal hydrophobic cleft of CaM, and conversely moves its hydrophobic face into an

exposed state. We predict that this apparently unfavourable position for melittin may be accommodated by a modified version of the Kretsinger bent helix model.

Table 5.1 shows that the polar face of the 14 residue amphiphilic segment of melittin is much less charged than mastoporan, containing only 1 positively charged lysine residue. Lys11 and Lys12 of mastoporan are substituted by Pro14 and Ala15 in melittin, both small, non-polar residues. The only other polar residues on the hydrophilic face are Thr10 and Thr11. Such small neutral polar residues are more amenable to burial at molecular interfaces because they can form satisfying hydrogen bonds with main chain carbonyl oxygen atoms in α helices (67). Therefore the movement of the polar melittin face into the C-terminal hydrophobic cleft may be reasonable.

Provided that the central helix of CaM could be distorted sufficiently, the exposed hydrophobic face of melittin could then be satisfied by interactions with apolar residues in the N-terminal hydrophobic pocket of CaM. By altering the original concept of Pereschini and Kretsinger (61) we have modelled a more severe kink of the central helix of CaM. The largest change again involved the conformation at Ser81 (ϕ = -85, ψ = 98). This alteration in torsion angle serves to bring the 2 hydrophobic domains of CaM sufficiently close to simultaneously bind a single peptide while still maintaining reasonable stereochemistry and hydrogen bonding patterns within the kinked central helix (see Fig. 5.6a). This mode of interaction would predict a more compact size for the CaM/melittin complex (~ 46 Å) than in native CaM (~ 65 Å) or that indicated in the Kretsinger CaM/MLCK model (~ 50 Å). Our model would also explain the 1:1 stoichiometry of the melittin/CaM complex, and the apparent involvement of both CaM domains in the binding of a single melittin peptide (52).

Detailed hydrophobic interactions of the CaM/melittin complex are shown in Fig. 5.6b. The majority of favourable nonpolar contacts involves residues in the N-terminal hydrophobic cleft of CaM; the primary residues involved (Phe19, Leu18, Leu36, Met51, Met71, Met72) all have analogous counterparts in the C-terminal domain and are the same residues implicated in the binding of mastoporan. Val5, Leu6, Val8, Leu9,

Leu13 and Ile20 of melittin form the significant hydrophobic face which binds into the CaM N-terminal hydrophobic cleft.

There are also several electrostatic interactions in the proposed melittin/CaM complex. Glu14 (CaM) is in a favourable position to interact with the positively charged N-terminus of melittin. Lys7 could undergo a favourable ionic interplay with Glu127. Lys21 is found lying near the carboxy terminus of helix F. Positively charged residues are often found at the C-terminal ends of helices where they interact with the negative component of the helix dipole (68). Arg22 could interact electrostatically with Asp78, Glu84 and Glu87. Lys23 caps the carboxy end of helix C. Arg24 is proximal to a cluster of acidic residues Glu47, Asp50 and Asp54 on helix C.

In addition to their role in charge-charge interactions, the aliphatic portions of Asp, Glu, Lys and Arg residues in CaM also serve to provide many hydrophobic interactions with melittin. For example, Arg74 and Lys75 of CaM surround and effectively bury the Trp side chain with the nonpolar portion of their side chains. This could explain the dramatic loss of fluorescence of Trp19 of melittin upon binding to CaM (66).

CONCLUSION

The main purpose of this study was to understand how amphiphilic helical molecules might associate with the calcium-binding regulatory proteins TnC and CaM. The experimental observation of a relatively favourable binding interaction of an α-helical segment into the C-terminal domain hydrophobic cleft (as seen in the high resolution picture of the TnC crystal lattice) was used as a guideline to model such a binding site. The modelled amphiphilic helices could be fit into an analogous position in the apolar cleft, forming several favourable molecular interactions without sacrificing standard stereochemical constraints. Their positions and interactions with various residues in the hydrophobic clefts which are exposed only in the Ca^{2+} bound state (11,19) would account for their Ca^{2+}-dependent affinity for CaM and TnC. In the Ca^{2+}-free form, the hydrophobic cleft would be substantially buried.

The interaction of extracellular peptides such as mastoporan and melittin with intracellular proteins such as TnC and CaM is not physiologically relevant. However because several of the binding sites for CaM on target molecules are thought to be relatively short amphiphilic helical regions our model may help in understanding the nature of these interactions.

ACKNOWLEDGEMENTS

N.C.J.S. thanks the Alberta Heritage Foundation for Medical Research for support. This research has been funded by a grant from the Medical Research Council of Canada to the Group in Protein Structure and Function at the University of Alberta.

REFERENCES

1. Kretsinger, R.H. 1980. Structure and evolution of calcium-modulated proteins. *CRC Crit. Rev. Biochem.* **8**:119-174.

2. Leavis, P.C. & Gergely, J. 1984. Thin filament proteins and thin filament-linked regulation of vertebrate muscle contraction. *CRC Crit. Rev. Biochem.* **16**:235-305.

3. Zot, A.S. & Potter, J.D. 1987. Structural aspects of troponin-tropomyosin regulation of skeletal muscle contraction. *Ann. Rev. Biophys. Biophys. Chem.* **16**:535-559.

4. Klee, C.B., Crouch, T.H. & Richman, P.G. 1980. Calmodulin *Ann. Rev. Biochem.* **49**:489-515.

5. Klee, C.B. & Vanaman, T.C. 1982. Calmodulin. *Adv. Protein Chem.* **35**:213-321.

6. Herzberg, O. & James, M.N.G. 1988. Refined crystal structure of troponin C from turkey skeletal muscle at 2.0 Å resolution. *J. Mol. Biol.* **203**:761-779.

7. Satyshur, K.A., Rao, S.T., Pysalska, D., Drendel, W., Greaser, M. & Sundralingam, M. 1988. Refined structure of chicken skeletal muscle troponin C in the two-calcium state at 2.0 Å resolution. *J. Biol. Chem.* **263**:1628-1647.

8. Babu, Y.S., Bugg, C.E. & Cook, W.J. 1988. Structure of calmodulin refined at 2.2 Å resolution. *J. Mol. Biol.* **204**:191-204.

9. Strynadka, N.C.J. & James, M.N.G. 1989. The crystal structures of the helix-loop-helix calcium-binding proteins. *Ann. Rev. Biochem.* **58**:951-998.

10. Herzberg, O., Hayakawa, K. & James, M.N.G. 1984. Crystallographic data for troponin C from turkey skeletal muscle. *J. Mol. Biol.* **172**:345-346.

11. Herzberg, O., Moult, J. & James, M.N.G. 1986. A model for Ca^{2+} induced conformational transition of troponin C and other calcium binding proteins. *J. Biol. Chem.* **261**:2638-2644.

12. Levine, B.A., Coffman, D.M.D. & Thornton, J.M. 1977. Calcium binding by troponin-C. A proton magnetic resonance study. *J. Mol. Biol.* **115**:743-760.

13. Laporte, P.C., Wierman, B.M. & Storm, D.R. 1980. Calcium-induced exposure of a hydrophobic surface on calmodulin. *Biochemistry* **19**:3814-3819.

14. Tanaka, T. & Hidaka, H. 1980. Hydrophobic regions function in calmodulin-enzyme interactions. *J. Biol. Chem.* **255**:11078-11080.

15. Cachia, P.J., Gariépy, J. & Hodges, R.S. 1985. *In* Calmodulin Antagonists and Cellular Physiology (Hidaka, H. & Hartshorne, D.J., eds.), Chapter 5, pp. 63-88, New York: Academic Press, Inc.

16. Drabikowski, W., Dalgarno, D.C., Levine, B.A., Gergely, J., Grabarek, Z. & Leavis, P.C. 1985. Solution conformation of the C-terminal domain of skeletal troponin C cation, trifluoperazine, and troponin I binding effects. *Eur. J. Biochem.*

151:17-28.

17. Dalgarno, D.C., Klevit, R.E., Levine, B.A., Scott, G.M.M., Williams, R.J.P., Gergely, J., Grabarek, Z., Leavis, P.C., Grand, R.J.A. & Drabikowski, W. 1984. The nature of trifluoperazine binding sites on calmodulin and troponin-C. *Biochim. Biophys. Acta* **791**:164-172.

18. Jarrett, H.W. 1984. The synthesis and reaction of a specific affinity label for the hydrophobic drug-binding domains of calmodulin. *J. Biol. Chem.* **259**:10136-10144.

19. Strynadka, N.C.J. & James, M.N.G. 1988. Two trifluoperazine-binding sites on calmodulin predicted from comparative molecular modelling with troponin-C. *Proteins: Structure, Function and Genetics* **3**:1-17.

20. Krebs, J. & Carafoli, E. 1982. Influence of Ca^{2+} and trifluoperazine on the structure of calmodulin. *Eur. J. Biochem.* **124**:619-627.

21. Guerini, D., Krebs, J. & Carafoli, E. 1987. Stimulation of the erythrocyte Ca^{2+}-ATPase and of bovine brain cyclic nucleotide phosphodiesterase. *Eur. J. Biochem.* **170**:35-42.

22. Klevit, R.E., Blumenthal, D.K., Wemmer, D.E. & Krebs, E.G. 1985. Interaction of calmodulin and a calmodulin-binding peptide from myosin light chain kinase: Major spectral changes in both occur as the result of complex formation. *Biochemistry* **24**:8152-8157.

23. Walsh, M., Stevens, F.C., Oikawa, K. & Kay, C.M. 1978. Chemical modification studies on the Ca^{2+}-dependent protein modulator: The role of methionine residues in the activation of cyclic nucleotide phosphodiesterase. *Biochemistry* **17**:3924-3930.

24. Thiry, P., Vandermeers, A., Vedermeers-Piret, M.-C., Rathe, J. & Christophe, J. 1980. The activation of brain adenylate cyclase and brain cyclic-nucleotide phosphodiesterase by seven calmodulin derivatives. *Eur. J. Biochem.* **103**:409-414.

25. Tanaka, T., Ohmura, T. & Hidaka, H. 1983. Calmodulin antagonists binding sites on calmodulin. *Pharmacology* **26**:249-257.

26. Prozialeck, W.C. & Weiss, B. 1982. Inhibition of calmodulin by phenothiazines and related drugs: Structure-activity relationships. *J. Pharmacol. Exper. Ther.* **222**:509-514.

27. Blumenthal, D.K., Takio, K., Edelman, A.N., Charbonneau, H., Titani, K., Walsh, K.A. & Krebs, E.G. 1985. Identification of the calmodulin binding domain of skeletal muscle myosin light chain kinase. *Proc. Natl. Acad. Sci. USA* **80**:3187-3191.

28. Lukas, T.J., Burgess, W.H., Predergast, F.G., Lau, W. & Watterson, D.M. 1986. Calmodulin binding domains: Characterization of a phosphorylation and calmodulin binding site from myosin light chain kinase. *Biochemistry* **25**:1458-1464.

29. Buschmeier, B., Meyer, H.E. & Mayr, G.W. 1987. Characterization of the calmodulin binding sites of muscle phosphofructokinase and comparison with known calmodulin-binding domains. *J. Biol. Chem.* **262**:9454-9462.

30. DeGrado, W.F., Erickson-Viitanen, S., Wolfe, H.R. & O'Neil, K.T. 1987. Predicted calmodulin-binding sequence in the γ-subunit of phosphorylase b kinase. *Proteins: Structure, Function, and Genetics* **2**:20-33.

31. Lin, C.R., Kapiloff, M.S., Durgerian, S., Tatemoto, K., Russo, A.F., Hanson, P., Schulman, H. & Rosenfeld, M.G. 1987. Molecular cloning of a brain-specific calcium/calmodulin-dependent protein kinase. *Proc. Natl. Acad. Sci. USA* **84**:5962-5968.

32. Erickson-Viitanen, S., O'Neil, K.T. & DeGrado, W.F. 1987. Theoretical and experimental approaches to the design of calmodulin-binding peptides: A model system for studying peptide/protein interactions. *In* Protein Engineering (Oxender, D.L. & Fox, C.F., eds.), pp. 201-211, New York: Alan R. Liss, Inc.

33. Hirai, Y., Yasuhara, T., Yoshida, U., Nakajima, T., Fujino, M. & Kitada, C. 1979. *Chem. Pharm. Bull.* **37**:1942-1949.

34. Haberman, V.E. & Jentsch, J. 1967. The amino acid sequence of bee venom melittin. *Hoppe-Seyler's Z. Physiol. Chem.* **348**:37-50.

35. Kabsch, W. & Sanders, C. 1983. Dictionary of protein secondary structure:pattern recognition of hydrogen bonded and geometrical features. *Biopolymers* **22**:2577-2637.

36. McCubbin, W.D., Oikawa, K. & Kay, C.M. 1986. Comparative calcium binding and conformational studies of turkey and rabbit skeletal troponin C. *FEBS Lett.* **195**:17-22.

37. Giedroc, D.P., Puett, D., Ling, N. & Stavos, J.V. 1983. Demonstration by covalent cross-linking of a specific interaction between β-endorphin and calmodulin. *J. Biol. Chem.* **258**:16-19.

38. Malencik, D.A. & Anderson, S.R. 1984. Peptide binding by calmodulin and its proteolytic fragments and by troponin C. *Biochemistry* **23**:2420-2428.

39. Malencik, D.A. & Anderson, S.R. 1982. Binding of simple peptides, hormones and neurotransmitters by calmodulin. *Biochemistry* **21**:3480-3486.

40. Malencik, D.A. & Anderson, S.R. 1983. Binding of hormones and neuropeptides by calmodulin. *Biochemistry* **22**:1995-2001.

41. Comte, M., Maulet, Y. & Cox, J.A. 1983. Ca^{2+}-dependent high affinity complex formation between calmodulin and melittin. *Biochem. J.* **209**:269-272.

42. Cox, A.J., Comte, M., Filton, J.E. & DeGrado, W.I. 1985. The interaction of calmodulin with amphiphilic peptides. *J. Biol. Chem.* **260**:2527-2534.

43. Malencik, D.A. & Anderson, S.R. 1983. High affinity binding of the mastoporans by calmodulin. *Biochem. Biophys. Res. Commun.* **114**:50-56.

44. Maulet, Y. & Cox, J.A. 1983. Structural changes in melittin and calmodulin upon complex formation and their modulation by calcium. *Biochemistry* **22**:5680-5686.

45. Sellinger-Barnette, M. & Weiss, B. 1982. *Mol. Pharmacol.* **21**:86-91.

46. Barnette, M.S., Daly, R. & Weiss, B. 1983. Inhibition of calmodulin activity by insect venom peptides. *Biochem. Pharmacol.* **32**:2929-2933.

47. Giedroc, D.R., Keravis, T.M., Stavos, J.V., Ling, N., Wells, J.N. & Puett, D. 1985. Functional properties of covalent β-endorphin peptide/calmodulin complexes. Chlorpromazine binding and phosphodiesterase activation. *Biochemistry* **24**:1203-1211.

48. DeGrado, W.F., Prendergast, F.G., Wolfe, H.R. Jr. & Cox, J.A. 1985. The design, synthesis, and characterization of amphiphilic inhibitory peptides of CaM. *J. Cell Biochem.* **29**:83-93.

49. Gledroc, P.P., Ling, N. & Puett, D. 1983. Identification of β-endorphin residues 19-25 as a region involved in the inhibition of calmodulin-stimulated phosphodiesterase activity. *Biochem.* **22**:5584-5591.

50. McDowel, L., Guatam, S. & Prendergast, F.G. 1985. Probable role of amphiphilicity in the binding of mastoporan to calmodulin. *Biochemistry* **24**:2979-2984.

51. Cachia, P.J., Van Eyk, J.E., Ingraham, R.E., McCubbin, W.D., Kay, C.M. & Hodges, R.S. 1986. Calmodulin and troponin C: A comparative study of the interaction of mastoporan and troponin I inhibitory peptide [104-115]. *Biochemistry* **25**:3553-3562.

52. Seeholzer, S.H., Cohn, M., Putkey, J.A., Means, A.R. & Cresi, H.L. 1986. NMR studies of a deuterated calmodulin with melittin. *Proc. Natl. Acad. Sci. USA* **83**:3634-3638.

53. Terwilliger, T.C. & Eisenberg, D. 1982. The structure of melittin. *J. Biol. Chem.* **257**:6016-6022.

54. O'Neil, K.T., Wolfe, H.R., Erickson-Vitanen, S. & DeGrado, W.F. 1987. Fluorescence properties of calmodulin-binding peptides reflect α-helical periodicity. *Science* **236**:1454-1456.

55. Collins, J.N., Potter, J.D., Norn, J.H., Wilshire, G. & Jackman, N. 1973. The amino acid sequence of rabbit skeletal muscle troponin C:gene replication and homology with calcium binding proteins from carp and hake muscle. *FEBS Lett.* **36**:268-272.

56. Wilkinson, J.M. 1976. The amino acid sequence of troponin C from chicken skeletal muscle. *FEBS Lett.* **70**:254-256.

57. Van Eerd, J.P. & Takahashi, K. 1975. Determination of the complete amino acid sequence of bovine cardiac troponin C. *Biochemistry* **15**:1171-1180.

58. Giedroc, D.P., Sinha, S.K., Brew, K. & Puett, D. 1985. Differential trace labelling of calmodulin: Investigation of binding sites and conformational states by individual lysine reactivities. *J. Biol. Chem.* **260**:13406-13413.

59. Linse, S., Drakenberg, T. & Forsén, S. 1986. Mastoporan binding induces a structural change affecting both the N-terminal and C-terminal domains of calmodulin A ^{113}Cd-NMR study. *FEBS Lett.* **199**:28-32.

60. Sanyal, G., Richard, L.M., Carraway III, K.L. & Puett, D. 1988. Binding of amphiphilic peptides to a carboxy-terminal tryptic fragment of calmodulin. *Biochemistry* **27**:6229-6236.

61. Persechini, A. & Kretsinger, R.N. 1988. Toward a model of the calmodulin-myosin light-chain kinase complex: Implications for calmodulin function. *J. Cardiovascular Pharm.* **12**:501-512.

62. Persechini, A. & Kretsinger, R.H. 1988. The central helix of calmodulin functions as a flexible tether. *J. Biol. Chem.* **263**:12175-12178.

63. Persechini, A., Blumenthal, D.K., Jarrett, H.W., Klee, C.B., Hardy, D.O. & Kretsinger, R.H. 1989. The effects of deletions in the central helix of calmodulin on enzyme activation and peptide binding. *J. Biol. Chem.* **264**:8052-8058.

64. Caday, G.G. & Steiner, S.F. 1986. The interaction of calmodulin with melittin. *Biochem. Biophys. Res. Commun.* **135**:419-425.

65. Steiner, R.F. & Norris, L. 1987. The interaction of melittin with troponin C. *Arch. Biochem. Biophys.* **254**:392-352.

66. Steiner, R.F., Marshall, L. & Needleman, D. 1986. The interaction of melittin with calmodulin and its tryptic fragments. *Arch. Biochem. Biophys.* **246**:286-300.

67. Argos, P. 1988. An investigation of protein subunit and domain interfaces. *Protein Engineering* **2**:101-113.

68. Richardson, J.S. & Richardson, D.C. 1988. Amino acid preferences for specific locations at the ends of α-helices. *Science* **240**:1648-1652.

Table 5.1

Alignment of TnC/CaM Inhibitory Peptides with the A-helix of TnC

A-helix	E16	E17	M18	I19	A20	E21	F22	K23	A24	A25	F26	D27	M28	-
Mastoporan	I1	N2	L3	K4	A5	L6	A7	A8	L9	A10	K11	K12	I13	L14
Mastoporan X	I1	N2	W3	K4	G5	I6	A7	A8	M9	A10	K11	K12	I13	L14
P. Mastoporan	V1	D2	W3	K4	K5	I6	G7	Q8	H9	J10	L11	S12	V13	L14
Melittin	A4	V5	L6	K7	V8	L9	T10	T11	G12	L13	P14	A15	L16*	I17

* The full sequence of bee venom melittin is G1 I2 G3 [A4 V5 L6 K7 V8 L9 T10 T11 G12 L13 P14 A15 L16] I17 S18 W19 I20 K21 K22 K23 K24 Q25 Q26; brackets indicate the peptide region we have modelled onto TnC and CaM.

Table 5.2

Summary of the Interactions of the Observed and Modelled Amphilic Helices with TnC*

TnC A-helix	E16	E17	M18	I19	A20	E21	F22	K23	A24	A25	F26	D27	M28
	E127 O197	D133			L122 V129 E127	O261			F105 L122 I121				I109 F105

Mastoporan	I1	N2	L3	K4	A5	L6	A7	A8	L9	A10	K11	K12	I13
	V129 E133 D136	E133 D136 D137	L137 V161	E127	V129 E127	L137 M152 M158 V161		T125	F105 I121	M158			I104 F105

Melittin	A4	V5	L6	K7	V8	L9	T10	T11	G12	L13	P14	A15	L16*
		L137 E133 D136	L137 V161	E127	L122 V129 E127	L137 M157 M158 V161				M158 F105			I104 F105

* Residues of the helices are given across the page in rows. Residues of TnC with which they interact are given below them in the columns. All distances are acceptable hydrophobic contacts (less than 3.8 Å) except for a few long contacts (4.4 Å or less) to alanine side chains.

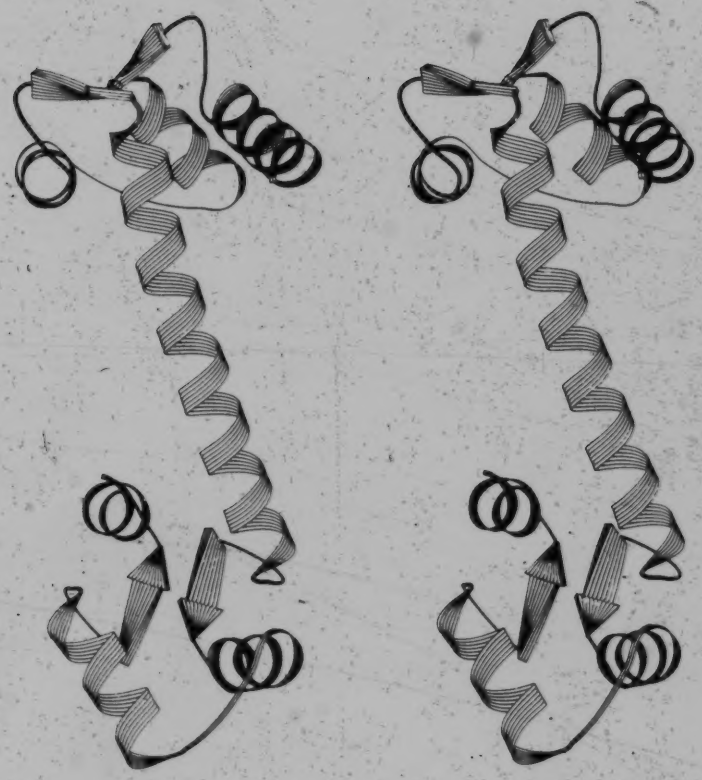
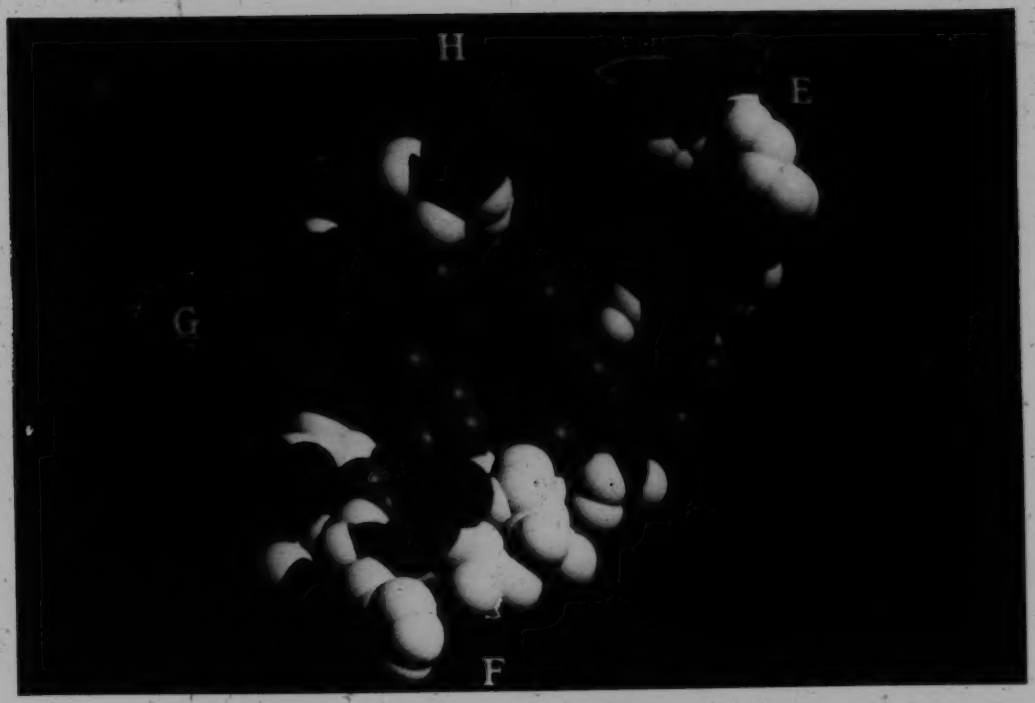
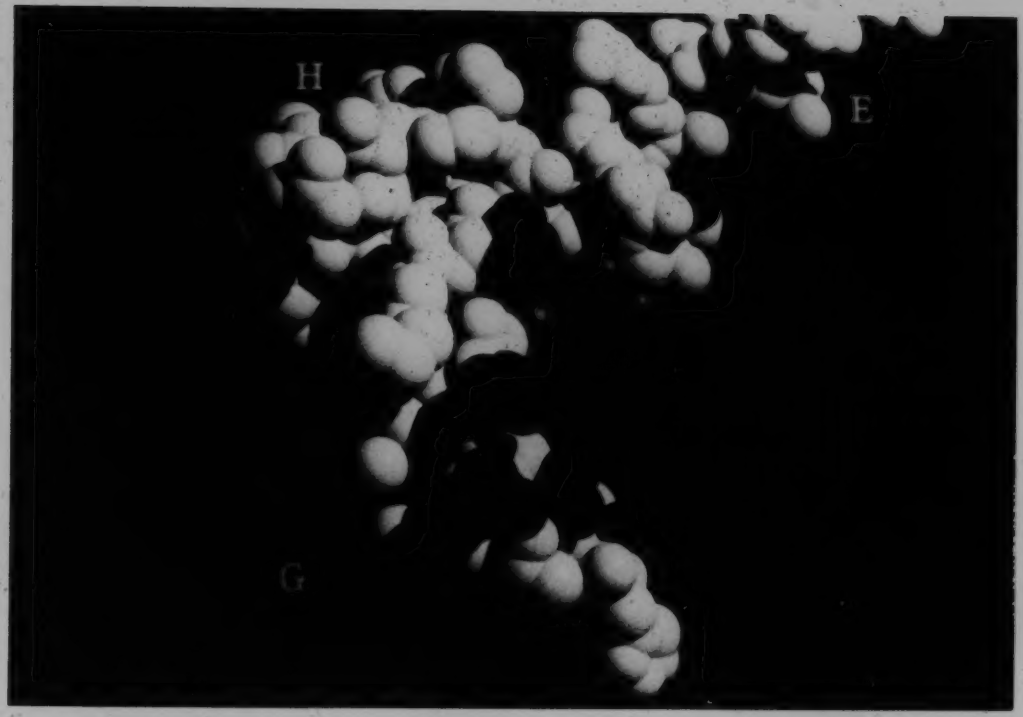


Fig. 5.1. A stereographic representation ribbon diagram of bovine brain CaM (8). This view shows the prominent secondary structural features that form the N and C-terminal Ca^{2+} binding domains and the long helical linker joining them.

a



b



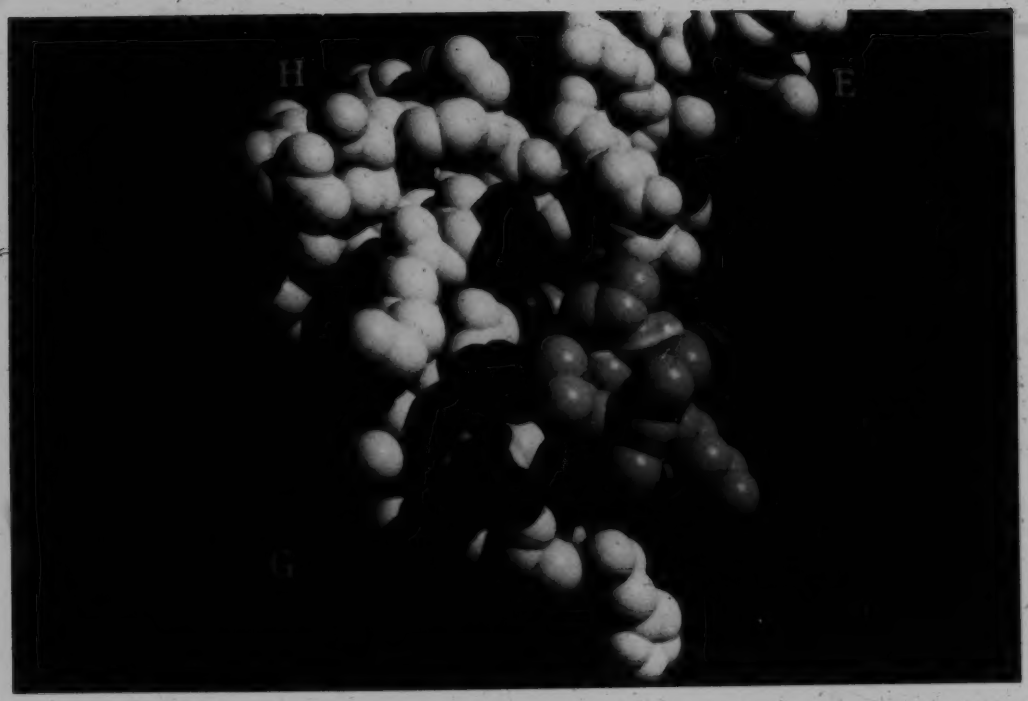
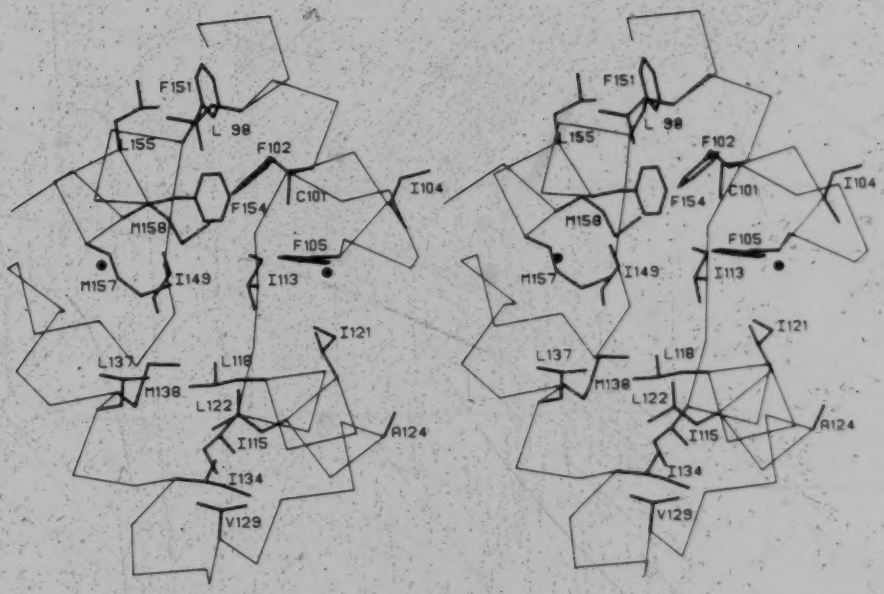


Fig. 5.2. a: A space-filling diagram of the C-terminus of turkey skeletal TnC looking directly into the C-terminal hydrophobic cup. Apolar residues are shown in green and acidic residues in red. Helices are labelled E, F, G and H according to the nomenclature in ref. 6.

b: A view 90° to that in 5.2a emphasizing the prominent hydrophobic cleft present in all the Ca^{2+} bound domains of TnC and CaM.

c: A space-filling representation of the proposed TnC-mastoporan complex. The C-terminal domain of TnC is shown in an identical orientation to that in 5.2b. The mastoporan helix that neatly fills the pronounced hydrophobic cleft is shown in yellow.

a


F151
L 98
L155
F102
C101
I104
F154
M158
F105
M157
I149
I113
I121
L137
L118
M138
L122
A124
I115
I134
V129
F151
L 98
L155
F102
C101
I104
F154
M158
F105
M157
I149
I113
I121
L137
L118
M138
L122
A124
I115
I134
V129

b

I9
A73
F12
M76
L69
F65
M72
A15
F68
M71
A57
L18
I63
F19
V55
I27
V35
I52
L32
L39
M51
V36
L48
A46
I9
A73
F12
M76
L69
F65
M72
A15
F68
M71
A57
L18
I63
F19
V55
I27
V35
I52
L32
L39
M51
V36
L48
A46

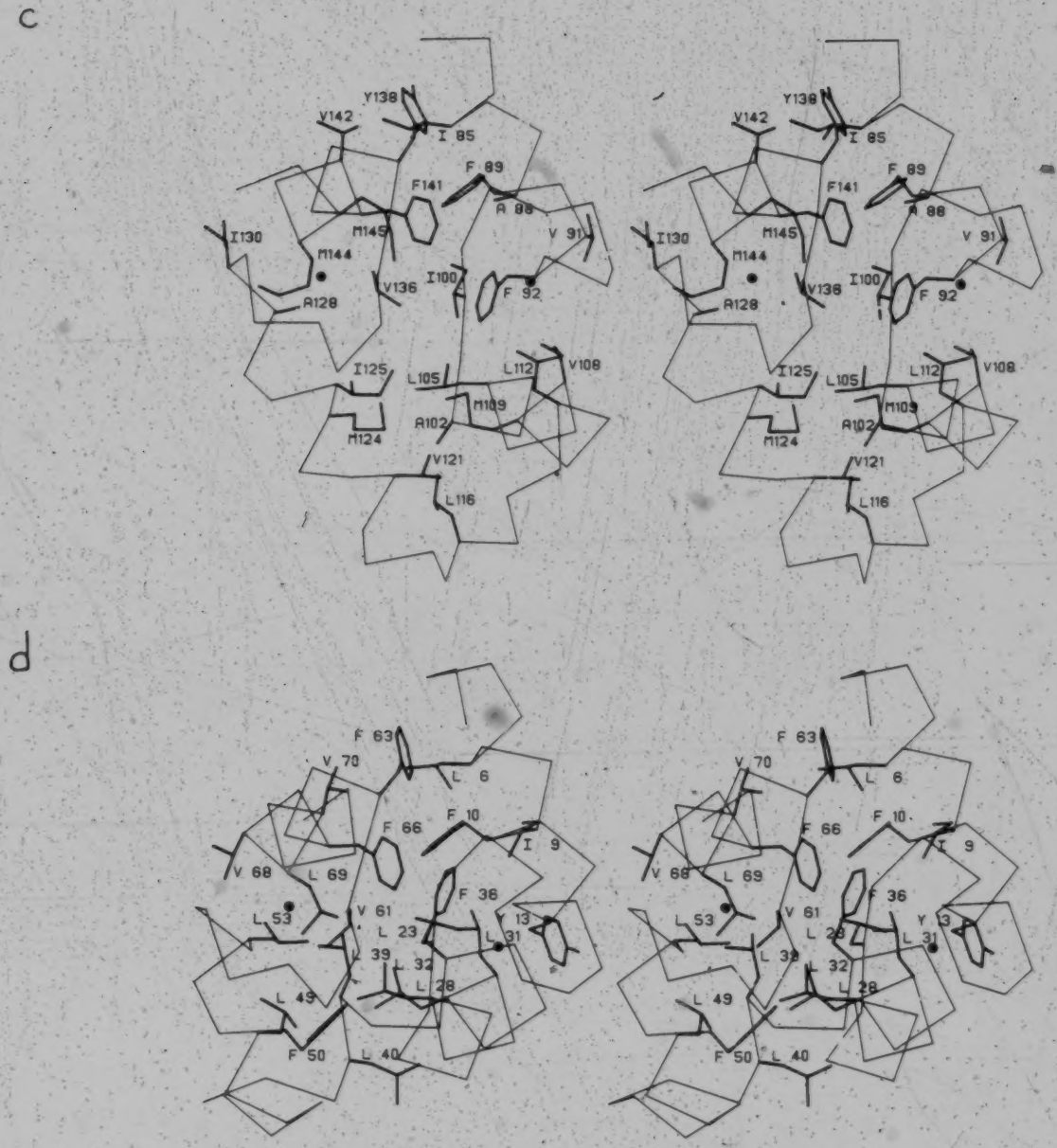


Fig. 5.3. Stereographic representations of the hydrophobic pockets of TnC and CaM. For clarity, only the hydrophobic side chains superimposed on the C^{α}-tracing are shown. Calciums are small black circles. All views are looking directly into the hydrophobic cup as in Fig. 5.2a.

a: the Ca^{2+}-bound C-terminal domain of TnC.
b: the Ca^{2+}-bound N-terminal domain of CaM.
c: the Ca^{2+}-bound C-terminal domain of CaM.
d: the Ca^{2+}-free N-terminal domain of TnC.

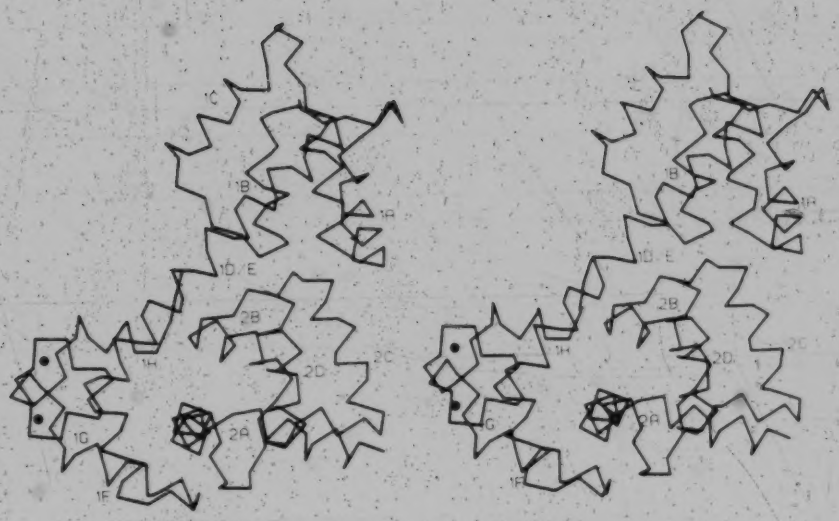
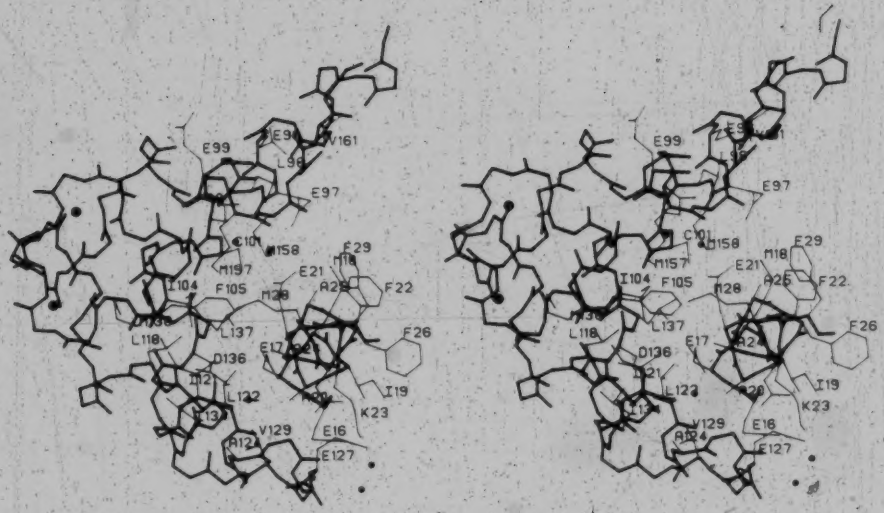
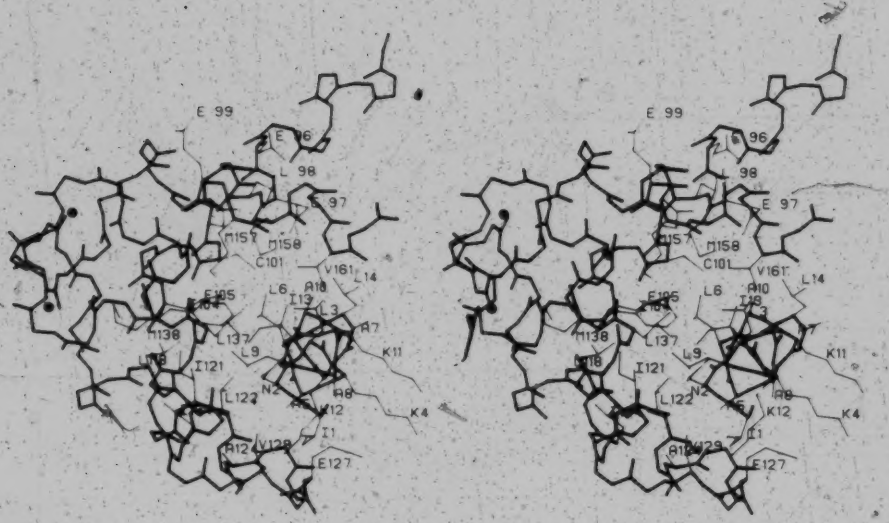


Fig. 5.4. A stereographic C^{α} tracing of the two symmetry related TnC molecules depicting their interaction as observed in the X-ray crystallographic structures (6,7). The helices of the two molecules are labelled 1A-1H, and 2A-2D/E [for nomenclature see (6)]. The C-terminal domain of molecule 1 is oriented as shown in Fig. 5.2b i.e. ~90° to the views in Fig. 5.2a and Fig. 5.3. Calcium ions are shown as black filled circles. The C-terminal domain of molecule 2 has been omitted to enhance the clarity.

a


E99
E97
C101
M158
M157
E21
I104
F105
F22
F26
L137
L118
D136
E17
I19
K23
L122
V129
E16
E127

b


E 99
L 98
E 97
M157
M158
C101
V161
L14
L6
I13
L3
A7
K11
M138
L137
L9
I121
L122
K12
K4
I1
E127

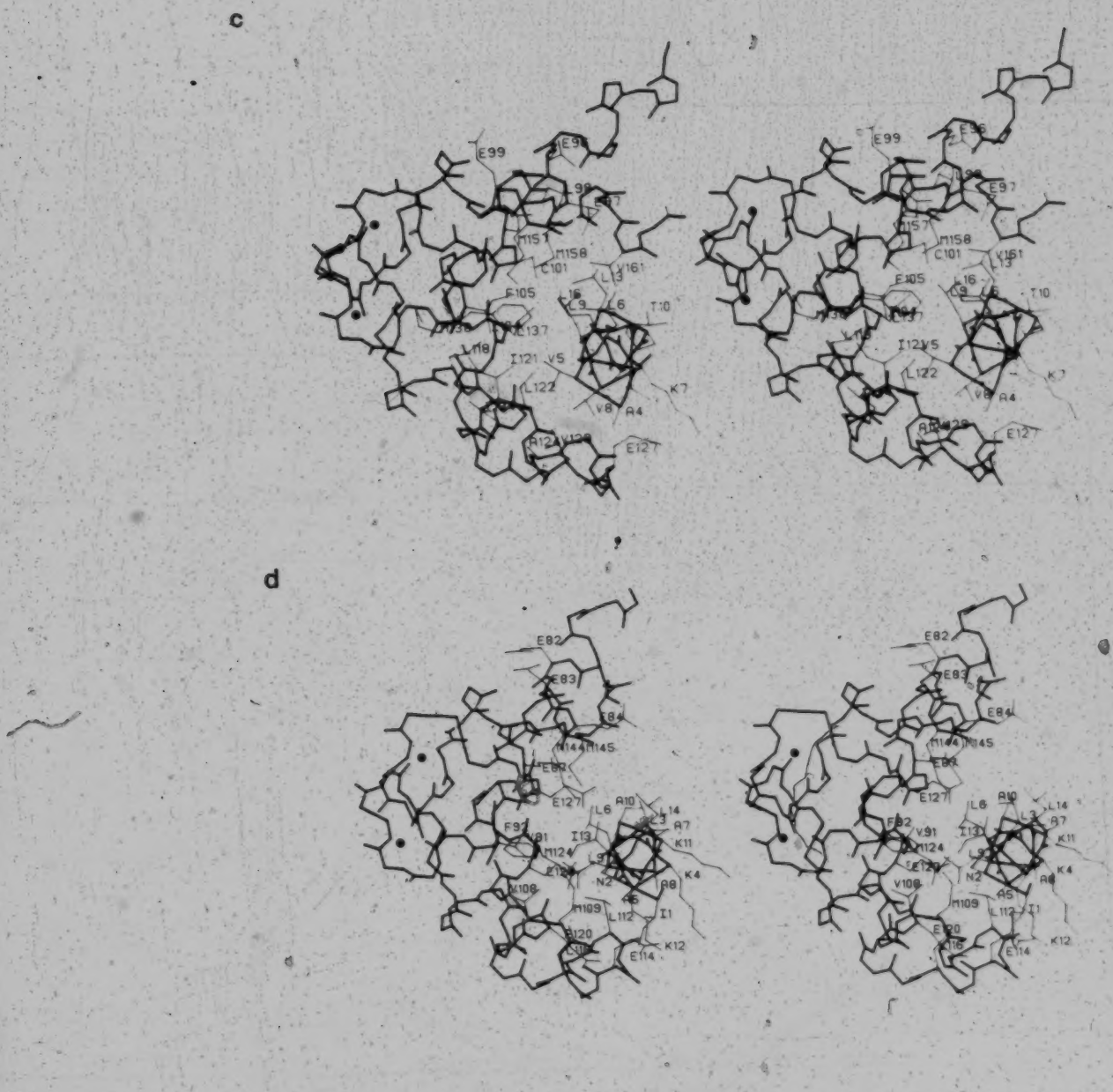
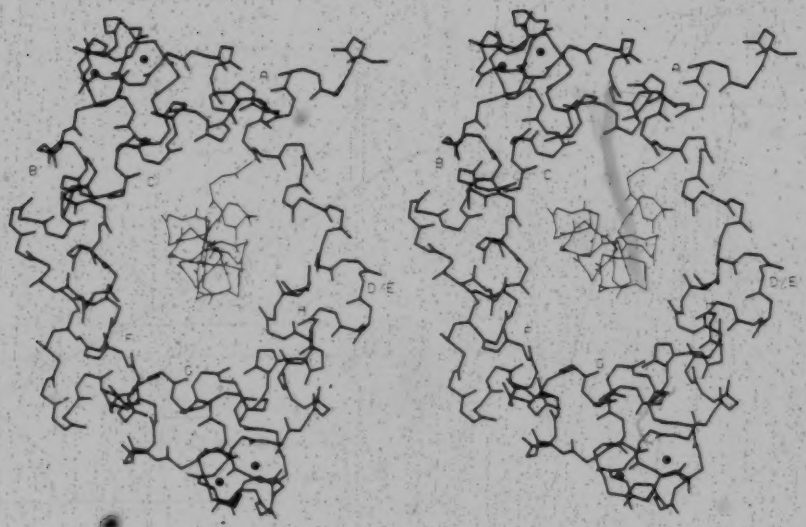


Fig. 5.5. The interactions of the C-terminal domain of TnC and CaM with amphiphilic helices: main chain is in dark lines, side chains in light. Calciums are black circles, waters are smaller black circles.

a: The crystallographically observed interaction of the TnC with the A-helic of its symmetry related neighbour.
b: the modelled interaction of mastoporan with TnC.
c: the modelled interaction of a segment of melittin with TnC.
d: the modelled interaction of mastoporan with the C-terminal domain of CaM.

a



b

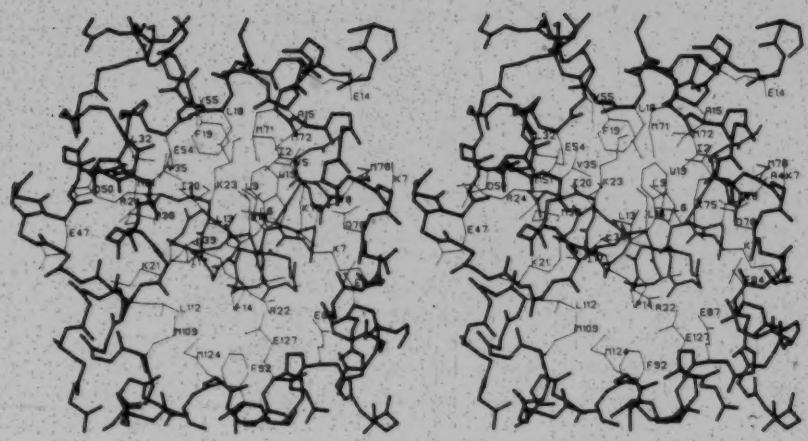


Fig. 5.6. a: A stereographic representation of the modelled interaction of melittin with bovine brain calmodulin. For clarity, only main chain atoms are represented. CaM is in thick lines, melittin in thin lines. Calciums are black circles. Helices of CaM are labelled A to H.

b: A stereographic view depicting side chain interactions within the melittin/CaM model. CaM and melittin main chain are in thick lines; side chains in thin lines. The orientation of the figure is identical to that shown in Fig. 5.5a, only enlarged.

CHAPTER 6

LYSOZYME REVISITED: CRYSTALLOGRAPHIC EVIDENCE FOR DISTORTION OF AN N-ACETYL-MURAMIC ACID RESIDUE BOUND IN SITE D[1]

ABBREVIATIONS

$(NAG)_n$, β(1→4) linked oligomer of 2-acetamido-2-deoxy-D-glucose; NAM, 2-acetamido-2-deoxy-D-muramic acid; NAL, δ-lactone of NAG; NAX, 2-acetamido-2-deoxy-D-xylose; GMGM, alternating β(1→4) linked oligosaccharide of NAG and NAM; MGM, β(1→4) linked trisaccharide NAM-NAG-NAM; GGG, β(1→4) linked trisaccharide NAG-NAG-NAG; HEWL, hen egg white lysozyme.

[1] A version of this chapter has been submitted for publication. Strynadka & James (1990). *J. Mol. Biol.*

ABSTRACT

A structure of the trisaccharide 2-acetamido-2-deoxy-D-muramic acid-β(1→4)-2-acetamido-2-deoxy-D-glucose-β(1→4)-2-acetamido-2-deoxy-D-muramic acid (NAM-NAG-NAM), bound to subsites B, C and D in the active site cleft of hen egg white lysozyme has been determined and refined at 1.5 Å resolution. The resulting atomic coordinates indicate that the NAM residue in site D is distorted from the full $^4C_1$ chair conformation to one in which the ring atoms C1, C2, O5 and C5 are approximately coplanar, and the hydroxymethyl group is positioned axially (a conformation best described as a sofa). This finding supports the original proposals that suggested the ground state conformation of the sugar bound in site D is strained to one which more closely resembles the geometry required for the oxocarbonium-ion transition state, the next step along the reaction pathway [Phillips, D.C., *Scientific American* 215:78-90 (1966)]. Additionally, detailed analysis at 1.5 Å resolution of the environments of the catalytic residues, Glu35 and Asp52, provides new information on the properties which may allow lysozyme to promote the stabilization of an unusually long-lived oxocarbonium-ion transition state.

Intermolecular interactions between the N-acetyl-muramic acid residue in site D and the lysozyme molecule that contribute to the saccharide ring distortion include: close packing of the O3' lactyl group with a hydrogen-bonded "platform" of enzyme residues (Asp52, Asn46, Asn59, Ser50 and Asp48), a close contact between the hydroxymethyl group of ring D and the 2'-acetamido group of ring C and a strong hydrogen-bonded interaction between the NH of Val109 and O6 of ring D which stabilizes the observed quasi-axial orientation of the -CH_2OH group. Additionally, the structure of this complex shows a strong hydrogen bond between the carboxyl group of Glu35 and the β-anomeric hydroxyl group of the NAM residue in site D. The hydrogen-bonded environment of Asp52 in the native enzyme and in the complex coupled with the very unfavorable direction of approach of the potential carboxylate nucleophile makes it most unlikely that there is a covalent glycosyl-enzyme intermediate on the hydrolysis pathway of hen-egg white lysozyme.

INTRODUCTION

Animal or C-type lysozymes are defined as 1',4-β-N-acetylmuramidases. They selectively cleave the glycosidic bond between the C1 of N-acetylmuramic acid (NAM) and the O4 of N-acetylglucosamine (NAG) in the bacterial cell wall peptidoglycan (Salton & Ghuysen, 1959, 1960; Jeanlos *et al.*, 1963, see Fig. 6.1). The best characterized member of the C-type lysozyme family is that isolated from hen egg white. The complete primary sequence of this molecule has been determined both by protein and by DNA sequencing techniques (Jollès & Jollès, 1961; Jollès, Jauragui-Adell, Bernier & Jollès, 1963; Canfield, 1963; Nguyen-Huu *et al.*, 1979). In common with all type C lysozymes, hen egg white lysozyme (HEWL) is comprised of 129 amino acid residues (Jollès & Jollès, 1984).

The tertiary structure of HEWL was initially determined at 2 Å resolution by X-ray crystallographic analysis of a tetragonal crystal form of the enzyme (Blake *et al.*, 1965, 1967a, 1967b; Phillips, 1966, 1967). Subsequently, the structure of HEWL has also been determined from tetragonal crystals under a variety of conditions of temperature and pressure (Kundrot & Richards, 1987) and from triclinic, monoclinic, and orthorhombic crystal forms (Joynson *et al.*, 1970; Moult *et al.*, 1976; Hogle *et al.*, 1981; Artymiuk *et al.*, 1982; Hodsdon, 1990). All of these studies indicate that the conformation of lysozyme is essentially the same under the different conditions and crystalline environments. The overall fold of lysozyme is shown in Fig. 6.2. A deep crevice or cleft containing the active site divides the molecule into two domains; one of them is almost entirely β-sheet structure (encompassing residues 40-85), whereas the other is comprised of the N and C-terminal segments (residues 1-39, and 101-129) and is more helical in nature. The two domains are linked by an α-helix (residues 89-99).

The currently accepted mechanism of lysozyme action, often termed general acid catalysis, was initially proposed from observations gleaned through the modelling of polysaccharide substrates into the active site cleft of the enzyme (Phillips, 1966, 1967; Blake *et al.*, 1967a, 1967b). These authors envisaged that six sugar units (termed A

through F) of an oligosaccharide substrate could be accommodated within the HEWL active site cleft (see Fig. 6.1). This proposal was supported by subsequent chemical data which showed that the natural bacterial cell wall substrate is indeed a hexasaccharide of alternating NAG and NAM residues. Additionally, the maximal rates of lysozyme action occur when either $(NAG\text{-}NAM)_3$ or the related chitin hexasaccharide $(NAG)_6$ are the substrates.

The modelled position of the hexasaccharide within the catalytic cleft led to the development of the lysozyme hydrolytic pathway as summarized in Fig. 6.3. Although several points concluded from these elegant analyses were later verified or supported by other experimental data (see Fig. 6.3), some aspects remained largely speculative. Most controversial of these are the proposed steps involving the ground state distortion of the site D NAM residue in order to avoid a steric clash with the side chain of tryptophan 108, and the electrostatic stabilization of an unusually long-lived oxocarbonium ion intermediate by aspartate 52.

An alternative mechanism for lysozyme action arises from the more general (e-e) glycosidase mechanism first proposed by Koshland (1953) and recently elaborated upon by Sinnott (1987). This mechanistic stream has been validated at all steps for a number of glycosidases including *E. coli* lacZ β-galactosidase and the β-glucosidase from *A. wentii* (Sinnott, 1987). Comparison of the Koshland and Phillips proposals indicates that the fundamental difference between the two mechanisms lies in whether there is stabilization of a long-lived carboxylate-oxocarbonium ion pair between Asp52 and the anomeric carbon atom of the NAM residue in site D (Phillips), or formation of a short-lived oxocarbonium transition state which could collapse into a tetrahedral covalent intermediate (Koshland). Early on, the modelling studies of Phillips and coworkers (1966) suggested that the formation of a covalent bond would be prohibited simply by the long distance from Asp52 to the C1 carbon atom of NAM (~ 3 Å). However, there has never been a high resolution, refined structural determination of a saccharide bound at site D to verify this claim. Furthermore, the existing experimental data describing the lysozyme hydrolytic pathway is

equally supportive of either mechanistic proposal (Sinnott, 1987). Unfortunately, the extremely difficult and time-consuming synthesis of lysozyme substrates and possible analogues in addition to the complications arising from transglycosylation effects (Kravachenko, 1967) and the formation of non-productive complexes (Holler *et al.*, 1975b) have severely limited the kinetic analysis required to definitively choose one mechanistic scheme over another.

In order to probe these controversial issues more fully, a trisaccharide (2-acetamido-2-deoxy-D-muramic acid)-β(1→4)-(2-acetamido-2-deoxy-D-glucosyl)-β(1→4)-(2-acetamido-2-deoxy-D-muramic acid), MGM, was co-crystallized with HEWL and the structure determined at 2.5 Å resolution (Kelly *et al.*, 1979). Due to the steric constraints imposed by the lactyl groups on O3 of the first and third sugar rings, MGM has only one strongly favored binding mode to HEWL in sites B, C and D (Patt *et al.*, 1978).

Although the above crystallographic work provided an exciting first glimpse of a NAM residue bound at the catalytic subsite D, the relatively low resolution and lack of refinement precluded an objective and accurate interpretation of the structure. Fig. 6.3 shows that subtle differences in atomic positions can either support or negate a proposed mechanistic event. For this reason we present in this paper a refined 1.5 Å analysis of the MGM-HEWL complex. In addition to a detailed profile of protein-carbohydrate interactions, the conformation of the enzyme-bound trisaccharide, the detailed environment of the catalytic residues Glu35 and Asp52 and the resulting implications on the catalytic mechanism of HEWL are discussed.

MATERIALS AND METHODS

Crystallization and Data Collection

Crystals of native HEWL (unpurified Sigma Lot 46F-86002) were grown at pH 4.2 using a concentration of 20 mg/ml protein in 5% NaCl buffered with 0.1 M NaOAc. The crystals of lysozyme with the MGM trisaccharide bound were grown by co-crystallization from a solution of 8 mg of MGM and 15 mg lysozyme in .5 ml of a 4% NaCl solution

buffered with 0.1 M NaOAc at pH 4.2. The native and MGM-bound crystals display remarkably different habits, thus giving an early indication that MGM had been incorporated into the tetragonal lysozyme unit cell in spite of the relatively weak dissociation constant K_d of 2.8 mM (pH 4.4, 25°C) (Patt *et al.*, 1978).

The unit cell dimensions of the native hen egg white lysozyme crystals (space group $P4_32_12$) are $a = b = 79.24$ Å and $c = 37.83$ Å, whereas those for the crystals of MGM bound lysozyme (again, space group $P4_32_12$) are $a = b = 78.88$ Å and $c = 38.34$ Å.

Intensity data for both native lysozyme and the lysozyme-MGM complex were collected using a S.D.M.S. twin area detector system (Hamlin, 1985) mounted on a Rigaku RU-200 rotating anode generator operating at 45 kV and 150 mA. The general data collection strategy was similar to that described previously (Xuong *et al.*, 1985). Data processing, including local scaling, Lorentz and polarization corrections, and merging to a unique data set were carried out with the program package of Howard *et al.* (1985). For native lysozyme a total of 60,233 observations were measured corresponding to a unique set of 11,235 reflections to 1.75 Å resolution. The overall merging R ($= \sum_j \sum_i | \bar{I}_j - I_{ji} | / \sum_j | \bar{I}_j |$) computed for all equivalent measurements was 0.044 and the estimated decay observed during the data collection was approximately 15%. For the lysozyme-MGM complex a total of 113,782 observations were measured corresponding to a unique set of 18,923 reflections to 1.5 Å resolution. Table 6.1 gives the breakdown of observed data with resolution. The overall merging R computed over all measurements was 0.054 and the estimated decay observed during the data collection was approximately 18%. The structure factor amplitudes for both the native and the enzyme-trisaccharide complex were brought to absolute scale using the scale factors determined from Wilson-type plots of the corrected intensity data (Wilson, 1942). The appropriately scaled data sets had an agreement factor of 0.12 ($\Sigma | | F_{MGM} | - | F_N | | / \Sigma | F_N |$, where $| F_{MGM} |$ and $| F_N |$ are the structure factor amplitudes measured from the MGM-lysozyme complex and native lysozyme crystals, respectively).

Refinement of Native Lysozyme

Restrained-parameter least-squares refinement (Hendrickson & Konnert, 1980) of the native HEWL structure was done using the data collected to 1.75 Å resolution (Table 6.2). The initial model used was set 2LYM from the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977) corresponding to the 1 atmosphere pressure refined structure of native lysozyme (Kundrot & Richards, 1987). The model consisted of 1,000 protein atoms and 19 of the water molecules having B-factors less than 25 $Å^2$. Minor adjustments to the initial model and the selection of additional solvent molecules were made from examining $|F_o|-|F_c|$ and $2|F_o|-|F_c|$ electron density maps computed at various stages during the refinement. The manual adjustments were made using FRODO (Jones, 1985) adapted to run on a Silicon Graphics Iris 3030 workstation (Cambillau, 1987).

During the least squares refinement, restraints on bond lengths, bond angles, planar groups and chiral centers for the protein were applied. Restraints on non-bonded contacts were weak and the weights on thermal parameter restraints were relaxed towards the final stages of refinement. No restraints on conformational angles or possible hydrogen bonds were applied at any stage.

Structure Solution and Refinement of the HEWL-MGM Complex

The lysozyme-MGM complex was solved by difference Fourier methods using the resulting fully-refined model of native lysozyme. Every precaution was taken to remove all solvent molecules from the active site region of the native lysozyme model before proceeding. From the remaining set, only water molecules with temperature factors, B, less than 25.0 $Å^2$ were included initially.

The refinement strategy for the sugar/enzyme data was similar to that described for the native lysozyme. A trisaccharide template was amalgamated into the PROLSQ dictionary using a modified version of CONEXN (Pahler & Hendrickson, 1990). The molecular template for the NAM and NAG residues in the trisaccharide was constructed

from data of the crystallographic determinations of the isolated sugar residues (Knox & Murthy, 1974; Mo & Jensen, 1978). In these structures the glucopyranose rings are in the full $^4C_1$ chair conformation. In order to ensure that the final refined conformation observed for the D ring was not biased, the following additional tests were done at the conclusion of the refinement. Firstly, the model of the N-acetyl muramic acid in site D was changed back to the full $^4C_1$ chair conformation with the hydroxymethyl C_6-O_6 group in an equatorial position and six further cycles of refinement were run. Secondly, with the site D NAM residue in the initial $^4C_1$ chair conformation, a water molecule was placed in the density that corresponded to the refined O6 position when it was quasiaxial. Six further refinement cycles with this as the 'initial' model were also carried out. Finally, the site D NAM residue was completely removed from the input coordinate set and five cycles of restrained least-squares refinement were run. $|F_o|-|F_c|$ and $2|F_o|-|F_c|$ maps were calculated with the resulting phases from all three test cases.

RESULTS

Quality of the Refined Structures

Forty-one cycles of refinement were carried out on native lysozyme and an additional 32 cycles on the lysozyme-MGM complex. Table 6.2 gives a summary of the refinement statistics for both the native and complexed structures. Fig. 6.4 depicts the main-chain conformational angles of the lysozyme component of the complex in the form of a Ramachandran plot (Ramakrishnan & Ramachandran, 1965). There are six non-glycine residues with ϕ,ψ conformational angles close to the left-handed α-helical region, α_L. All the other amino acids in the region with ϕ from + 50° to + 120° correspond to glycine residues. There are no other residues in the refined HEWL with conformations corresponding to unallowed regions of ϕ,ψ space.

Three of the residues with ϕ,ψ values close to the α_L region are amino acid types seldom found with positive ϕ values, Arg21, Phe38 and Glu57. Arg21 occurs in two

successive β-turns, Asn19→Gly22 and Tyr20→Tyr23. The first is a type II* turn and the second is a type I'. Phe38 is the third residue in a type III' turn that runs from Ser36 to Asn39. Gln57 is the last residue in a type I turn; this turn is buried behind the catalytic residue Glu35. Gln57 is probably involved in substrate binding interactions in site E.

The average temperature factor for the atoms of native lysozyme is 18.3 $Å^2$, and for the atoms of lysozyme in the sugar complex it is 15.4 $Å^2$. Fig. 6.5 compares the main-chain and side-chain B factors between lysozyme and the lysozyme-trisaccharide complex. As seen in Fig. 6.5 and from the lower average temperature factor of the complex, the whole lysozyme molecule becomes slightly more ordered upon sugar binding. The atoms exhibiting the greatest decreases in thermal motion are those from residues Arg68 to Arg73 and generally those which line the active site cleft. Earlier molecular dynamic trajectory simulations of free and substrate bound lysozyme had concluded that a region of the active site (Asp101 to Ala107) had significantly reduced motion, presumably due to interactions with the substrate (Post *et al.*, 1986, 1989).

In general, the most poorly ordered residues of HEWL in the native and complexed structures involve the side chains of lysine, arginine, or glutamine residues which lie on the molecular surface. In the native structure the side chains of Lys13, Arg14, Arg21, Arg61, Arg73, Lys97, Asp101, Arg112, Gln121, Arg125, Arg128 and the C-terminal residue Leu129 are not well defined (temperature factor's of over 35 $Å^2$; weak electron density for the atoms at the end of the side chains). The side chains of Arg21, Arg73, Gln121, Arg125, Arg128, and Leu129 are similarly affected in the trisaccharide bound structure.

Alternative conformations of the side chain position were observed in the final electron density for Val109 in both the native and the complex protein structures. These alternative conformations correspond to side chain torsion angles of 65°, -55° and 178°. In the native structure Val109 is exposed to solvent; none of the three possible modes appears to have an energetic advantage over the other. Thus, the equivalence of the three

---

* Turns have been classified following Crawford *et al.* (1973).

lobes of electron density, suggesting a hindered-rotor like motion of the side chain, may not be surprising. In the case of the trisaccharide-lysozyme complex, Val109 can form hydrophobic contacts with the D ring only in the 65° or -55° conformations. In both of these cases, one of the methyl groups makes contact with the site D sugar and the other is solvent exposed. There is no significant statistical difference or better fit of electron density between the two possible conformations upon refinement so we assume equal occupancy for each.

In order to substantiate further discussion on enzymatic pathways and details of the interactions between HEWL and MGM, it is important to verify the quality of the structures on which these discussions are based. The observed electron densities for the side chains of residues in the active site region of lysozyme in the native and complexed form are shown in Fig. 6.6a and Fig. 6.6b, respectively. As surmised from these figures, all atoms of the proposed catalytic residues, Glu35 and Asp52, are clearly defined in both structures. Furthermore, Asp46, Asp48, Gln57, Asn59, Trp62, Trp63, Asp101, Asn103 and Ala107, all residues which directly contact the trisaccharide MGM, are also clearly observed in the electron density maps. The lower B factors (Fig. 6.5) and enhanced electron densities for Asp101 and Trp62 (Fig. 6.6) indicate increased order for these residues in the liganded form of lysozyme.

There is no satisfactory way of determining the accuracy of the atomic coordinates resulting from restrained-parameter least-squares refinement. We felt that the approach of Cruickshank (1949, 1954, 1967), which estimates the coordinate error of each atom type as a function of its temperature factor, would be particularly appropriate in the present work as it emphasizes the fact that coordinate errors in regions of low B-factors are small, whereas the coordinates of those portions of the molecules or solvent with high B factors are less reliable (see also Read *et al.*, 1983; Fujinaga *et al.*, 1985). The calculated root mean square errors for the protein atoms and associated solvent in the native lysozyme structure is 0.16 Å, and in the lysozyme-MGM complex is 0.14 Å. In a more general

sense, though, the coordinate accuracy of a particular atom can be estimated from its observed electron density and refined individual temperature factors. The observed electron density for each of the atoms of MGM and the corresponding group average temperature factors are given in Fig. 6.7 and Table 6.3, respectively. The relatively low temperature factors and clearly defined electron density associated with the atoms of ring C show that this is the most highly ordered of the three rings, with the buried 2-acetamido group exhibiting B factors that are as low as some of the most well defined groups in the protein. The atomic positions of the site C NAG are thus determined with a high accuracy. On the other hand, the solvent exposed 2-acetamido group of ring D and the lactyl group of ring B are poorly defined in the electron density map and have the largest B factors of all the atoms in the MGM trisaccharide (Table 6.3). The atomic positions of those groups are thus determined with lower accuracy. The acetamido group of ring B and the lactyl group of ring D, although better defined than the site B lactyl and site D N-acetamido groups (Fig. 6.7b), also have relatively large B factors (Table 6.3). The atomic positions for these groups are thus less accurate than those of the corresponding groups on ring C.

Comparison of HEWL in the Native and Complexed Form

PROTEIN ATOMS. The native lysozyme atomic coordinates were compared with those of the lysozyme molecule in the structure of the complex using a least-squares fitting program (W. Bennett, unpublished). The r.m.s. deviation for the 1,000 protein atoms is 0.79 Å; for the 517 main-chain atoms the r.m.s. difference is 0.28 Å. The overlap of the two lysozyme structures is shown in Fig. 6.8. In general, the changes in the molecular conformation are subtle and involve a narrowing of the cleft as the residues and the secondary structural units that line the sugar binding site move in towards the bound trisaccharide. This result correlates well to an earlier theoretical prediction employing normal mode analysis which concluded that the motions of lysozyme associated with the low frequency modes (the dominant motions) involved a slight opening and closing of the active site cleft via concerted movements of the flanking structural regions (Levitt *et al.*,

1985).

The three specific regions that exhibit the greatest conformational differences between the two lysozyme molecules are Asn59 to Cys64, Val99 to Thr118, and Arg68 to Cys76. The first two of these regions contribute directly to carbohydrate binding. Amino acids Asp52, Trp62, Trp63, Asp101, Asn103 and Val109, contained within these regions, all shift in position to accommodate the incoming sugar (Fig. 6.8, see below for specific contacts). The third region mentioned above packs against the carbohydrate binding loop Asn59 to Cys64. It shifts in position as well as becoming more ordered (see Fig. 6.5), even though it does not directly make contacts with the carbohydrate substrate. Conformational shifts in this region of the lysozyme molecule have been observed previously in the complexes of several N-acetamido monosaccharides (Perkins *et al.*, 1978), and of $(NAG)_3$ NAL binding to HEWL (Ford *et al.*, 1974). In the present work the conformational change observed for the Arg68 to Cys76 region is also accompanied by a change in the hydrogen bonding scheme to the Asn59 to Cys64 loop. The key residues in this hydrogen bonding network are Asn59, Ser60, Arg61, Thr69, Ser72, and Arg73. One dramatic result of the change in hydrogen bonding is a flip, in the complexed molecule, of the peptide bond between Arg73 and Asn74. Flexibility in this peptide was predicted from the calculated atomic fluctuations derived from molecular dynamics simulations (Post *et al.*, 1986).

SOLVENT MOLECULES. An overall comparison of the distribution of solvent molecules between the two structures determined in this work shows 23 water molecules with coinciding atomic positions (.2 Å or less). An additional 38 water molecules are within 0.2 to 0.5 Å apart in position and another 11 within 0.5 to 1.0 Å apart.

Analysis of the solvent molecules observed in the active site clefts of the native and complexed forms of lysozyme indicate that the binding of the trisaccharide displaces seven ordered water molecules from the active site region. Three of these waters are displaced by the C ring and four by the D ring. The binding site for the B sugar ring does not appear to

have any strongly ordered water molecules in the native structure that are displaced by MGM. Interestingly, six of the water molecules that are displaced from the active site cleft of the native molecule lie in positions analogous to those occupied by atoms of the MGM carbohydrate in the complex structure. These atoms (with distances from the displaced water molecules given in parentheses) include O1 (.41 Å), O4 (.97 Å), C5 (.87 Å), and O6 (.35 Å) of the D sugar unit and O7 (.39 Å) and C8 (.74 Å) of the C sugar unit.

Structural Features of the HEWL-MGM Complex

MGM binds to the active site cleft of HEWL in an extended conformation (Figs. 6.8 and 6.9). The geometry of the glycosidic linkages between the monosaccharides is characteristic of β(1→4) linked sugars (Table 6.4). The junctions between the two saccharide units in sites B and C of MGM and of the trisaccharide NAG-NAG-NAG (GGG) bound to HEWL (Cheetham *et al.*, 1990) are virtually identical. There are also intramolecular hydrogen bonds between O3' of the sugar in site C and the ring oxygen of the site B sugar in both trisaccharides. The very different values for ϕ describing the glycosidic linkages A-B for GGG and C-D for MGM are outside the range observed for this parameter (-71° to -96°) in other oligosaccharide crystal structures (Mo & Jensen, 1978; Jeffrey, 1990). The wide range of the helical-twist parameter, ψ_H in Table 6.4, is evidence of considerable flexibility in the β(1→4) linkage. Values of ψ_H that are far from 0° preclude the O5...H-O3' hydrogen bond that is characteristic of the β(1→4) linkage.

SITE B N-ACETYLMURAMIC ACID. The NAM residue bound in site B has the full $^4C_1$ chair conformation of the sugar ring (Fig. 6.10a). Bond lengths, valence angles and torsion angles for the atoms of this saccharide fall within the ranges for those equivalent parameters summarized by Jeffrey & Taylor (1980). The N-acetyl group on C2 and the hydroxymethyl group on C5 are well defined in the electron density map (Fig. 6.8b). The plane of the N-acetyl group is approximately perpendicular to the general plane through the 6 atoms of the pyranose ring. The torsional angle C1-C2-N2-C7 is 114°. The lactyl group

of NAM in site B is poorly defined in the electron density map. It extends into the bulk solvent region of the crystal. The lactyl group conformation was chosen to place the carboxylate into significant density (C2-C3-O3-C9, 89°; C3-O3-C9-C10, -32°). In this conformation the lactyl carboxylate makes a hydrogen bond to Asn103 $N^{\delta 2}$ (Fig. 6.10a). Other hydrogen bonded interactions that involve the NAM residue in site B are from the Asp101 carboxylate to O6 and from O6 and the carbonyl oxygen of the N-acetyl group to a group of three hydrogen-bonded waters (O236, O270 and O163) that link the sugar to the binding surface of HEWL (Fig. 6.10a and Table 6.5).

The other major binding interaction for site B is the extensive non-polar contact between the indole ring of Trp62 and the non-polar surface of the NAM residue (Fig. 6.10a). The planes of these two rings are approximately parallel and separated by ~3.8 Å on average. The stacking of aromatic side chains on to the sugar rings of protein-bound mono saccharides has been observed previously in both the high resolution structural determinations of L-arabinose binding protein and D-galactose binding protein (Quiocho, 1988, 1989). It has been proposed that this type of interaction may be a general feature of protein-saccharide binding (Quiocho, 1986).

SITE C N-ACETYLGLUCOSAMINE. Site C plays a key role in determining the cleavage specificity of HEWL (Fig. 6.10b). There is a deep pocket on the surface of HEWL into which the 2-acetamido group fits (Fig. 6.9). This determines which side of the sugar ring faces the enzyme, since any substituent on O3 is precluded on steric grounds. Thus, only N-acetylglucosamine residues will fit into site C (Phillips, 1966; Blake *et al.*, 1967b; Johnson *et al.*, 1988).

The ring of the N-acetylglucosamine adopts the full $^4C_1$ chair conformation as was observed for the NAM residue in site B. Similar conformations for both of these rings have been observed in the high resolution, refined structure of a $(NAG)_3$ complex with HEWL (Cheetham *et al.*, 1990). The plane of the N-acetyl group is approximately perpendicular to the mean plane of the pyranose ring as was observed for NAM in site B.

The conformational angle C1-C2-N2-C7 is 115°.

The 2-acetamido moiety of the NAG ring in site C forms two hydrogen bonds with main chain atoms of HEWL, from N2 to the carbonyl oxygen of Ala107 and from the main chain NH of Asn59 to the carbonyl oxygen O7 (Table 6.5). In addition, the N^{ε} atom of Trp63 donates a hydrogen bond to O3 of the ring and the N^{ε} atom of Trp62 donates a hydrogen bond to O6. Hydrophobic contacts are made from the acetamido group on C2 to the indole ring of Trp108 and the side chains of Ile58 and Ile98 (Fig. 6.10b).

The well-defined electron density (Fig. 6.7b) and the low B-factors for the NAG ring in site C and its attached N-acetyl group (Table 6.3) confirm that the occupancy of MGM in these crystals is almost 1.0. The structure of the complex has been refined on this basis since only the B-factors of the MGM molecule were varied; the occupancy factors for all atoms of MGM were kept at 1.0.

The carbohydrate/protein interactions we observe for the NAG residue bound to subsite C of HEWL are similar to those found in previously determined complexes with a saccharide bound at subsite C (Perkins *et al.*, 1978; Cheetham, 1990). That this residue is very important to the substrate binding interactions is seen from the estimate of the free energy of binding for this site (approximately -24 kjoules), a value almost twice that of any of the other 5 individual saccharide binding sites (Chipman & Sharon, 1969).

SITE D N-ACETYL MURAMIC ACID. The NAM residue that binds to site D is distorted from the expected $^4C_1$ chair conformation (Fig. 6.10c). It most closely approaches a sofa conformation with C2, C1, O5, C5 and C4 approximately coplanar (Fig. 6.11). Apart from the large distortion towards a more axial position for C6-O6, the amount of distortion of the atoms in the pyranose ring is not large and borders on the level of accuracy of the present structure determination. Therefore, the tests that were carried out following the completion of the refinement (see section (c) of Materials and Methods) were designed to confirm this observed distortion. From Fig. 6.10c it can be seen that the NAM residue in site D packs between residues Trp108, Val109 from the "top" and Asn46, Asp52

and Asn59 from "below". This tight packing has rather severe restrictions on the conformations of the lactyl group on O3 and on the position of the hydroxymethyl group, C6-O6. The lactyl group packs against the side chains of Asn46 and Asn59 thereby being forced to make close contacts with O5 and C6 of the NAG residue in site C. The hydroxymethyl group C6-O6 is in a quasi-axial orientation. This results from relieving the possibly too-close contacts with the 2-acetamido group on the NAG ring in site C. Based on the results of model building experiments Phillips (1966) had suggested that this conformation was due to close contacts with Trp108 and Glu35. In fact, the present refined structure shows that these residues are far from the NAM residue in site D and do not contribute to the distortion. As shown in Table 6.3 the atoms of the 2-acetamido group on NAG have very low B-factors, and are as well ordered and tightly bound as the surrounding protein atoms. An equatorial orientation for C6-O6 on ring D would make serious too-close contacts with the highly ordered acetamido N2 of ring C and so the hydroxymethyl group is forced into the observed quasi-axial position.

The first refinement test (section (c) of Materials and Methods) confirmed that starting from an initial full $^4C_1$ chair conformation 6 cycles of least-squares refinement resulted once again in a distorted sofa conformation for the NAM ring. The C6 and O6 atoms in the equatorial starting positions were 1.0 and 1.4 Å, respectively, from the final refined quasi-axial positions. In the second refinement test, a water molecule was placed in the density that corresponded to the axial O6 atom in addition to assuming that the full $^4C_1$ chair conformation with C6-O6 equatorial was once again the initial conformation of the NAM residue. The electron density in the region of the site D NAM saccharide resulting from the additional six cycles of least-squares refinement was virtually indistinguishable from that after the first test and from the final $2|F_o|-|F_c|$ electron density distribution resulting from the main body of refinement. In this second test however, the equatorial C6-O6 hydroxymethyl group did not move into its former density; the extra water molecule remained in its starting position and no new density was observed for the C6-O6 atoms in

an equatorial position. Finally, Fig. 6.12 shows the $|F_o|-|F_c|$ electron density resulting from the third test, i.e. removal of the entire D sugar from the input model followed by five cycles of least-squares refinement. The difference density observed at site D clearly supports the distorted conformation obtained from our initial refinement and indicates that this result is not a manifestation of model bias. Collectively, the results of these three refinement tests confirm that the sugar ring in site D is distorted in the ground state towards the sofa or half chain conformation.

Hydrogen bonding interactions from the enzyme to the NAM residue in site D help to maintain the distortion (Table 6.5). There is a strong hydrogen bond from the main chain NH of Val109 to O6 of the quasi-axial hydroxymethyl group. This is a favorable interaction that assists the distortion of the D ring. It would not be made if the NAM residue had the full $^4C_1$ chair conformation.

The importance of this hydrogen bond is underscored by the results of two independent studies. Ballardie *et al.* (1977) tested a series of aryl β-1,4-linked oligosaccharides of $(NAG)_n$, with n = 2, 3 and 4 as substrates of HEWL. Kinetic parameters and the pH dependence of these substrate analogues paralleled the kinetic behaviour of $(NAG)_6$. However, compounds of the type $(NAG)_{2-4}$ (NAX)-DNP (DNP is dinitrophenol) in which the 2-acetamido 2-deoxy-β-D-xylose lacks the hydroxymethyl group on C5 were not hydrolyzed by lysozyme. Even though the equivalent oligosaccharide $(NAG)_3NAX$ binds tightly to HEWL with a K_d of 4.3×10^{-6} (Schindler *et al.*, 1977), the NAX derivatives are poor substrates. Recently, it has been reported that a lysozyme mutant in which Val109 was substituted by a proline residue has markedly reduced hydrolytic activity towards oligosaccharide substrates (Inaka *et al.*, 1990). The cyclic nature of the imino-peptide bond (Trp108-Pro109) in this mutant precluded the stabilizing hydrogen-bonding to the O6 atom of the saccharide in site D. The presence of a prolyl residue at position 109 may also seriously destabilize (and even preclude) productive binding of NAG or NAM residues in site D (Fig. 6. 10c). An interesting experiment with

such a mutant would be to test its ability to hydrolyze the $(NAG)_n$ (NAX)-PNP substrates of Ballardie *et al.* (1977).

A second important sugar-protein hydrogen bond at site D is that which the glycosidic oxygen O1 on the anomeric carbon forms with the carboxylate of Glu35. This interaction favors the binding of the β-anomer of the reducing sugar over that of the α-anomer. Furthermore, examination of the structure shows that there would be no room for an α-anomeric oxygen on this ring (Fig. 6.10c). Thus, there is no crystallographic evidence of an α-D-glucopyranose configuration for the NAM residue in site D. This clearly contradicts the NMR results of the measurement of a coupling constant H_1^{α} H_2 for NAM in site D reported by Patt *et al.* (1978).

Fig. 6.13 shows a superposition of the N-acetylmuramic acid residue from site D with the structure of the corresponding uncomplexed monosaccharide (Knox & Murthy, 1974). The two molecules were superimposed by overlaying the C2, N2, C3, O3, C4 and O4 atoms from each structure. The atoms exhibiting the largest differences are O5, C5, C6, O6 and the atoms of the lactyl group on O3. In addition, since the structure of the monosaccharide corresponds to that of the α-anomer, the anomeric oxygen atoms (O1) positions are very different. In the monosaccharide the orientation of the acetamido group favors a hydrogen bond to the carboxylate of the lactyl group. It is not possible to form this hydrogen bond in the NAM residue bound to site D since the lactyl group moves up between ring C and ring D to avoid a too close contact with Asn46 and Asn59 in the complex. The shift of C6-O6 into the quasi-axial position is the result of the too close contact with the 2-acetamido group on NAG in site C in the full chair equatorial form discussed above.

Comparison of the Refined HEWL-MGM with Previous Complexes of HEWL

There are two previously published crystal structure determinations of oligosaccharides bound to HEWL in which site D is occupied (Ford *et al.*, 1974; Kelly *et al.*, 1979). Unfortunately, both of these structures are at relatively low resolution (2.5 Å) and

are not refined. In general, the interpretation of the detailed binding interactions made by small molecules to proteins is difficult when done from difference electron-density maps. Previous studies from this laboratory, on the binding of inhibitor molecules to aspartyl proteinases (James, Sielecki & Moult, 1983; James & Sielecki, 1987), have shown that upon least-squares refinement of the complex there is often a shift in position of the inhibitor (up to 1.0 Å) from that deduced from the difference maps. This was the case even when the quality of the initial difference electron density was exceptionally good. It appears from the comparison of the structure of MGM for the unrefined HEWL-MGM complex (Kelly *et al.*, 1979) to the present structure that the position of the difference density in that earlier study is shifted by approximately 1.0 Å out of the cleft and towards the solvent (Fig. 6.14a). This has affected the interpretation of the binding interactions. As well, the lower resolution of the data used for that work precluded the detection of the distortion of the NAM ring bound in site D.

Differences in the enzyme-carbohydrate interactions in sites B, C and D between the present refined HEWL-MGM complex and the HEWL-$(NAG)_3$-NAL (the δ-lactone of NAG_4, Ford *et al.*, 1974) are also pronounced (Fig. 6.14b). It is not possible to tell whether these differences are real or whether they are due to the low resolution and/or lack of refinement of the crystallographic structure of the δ-lactone tetrasaccharide. Again the displacement of common atoms in these two structures is of the order of 1.0-1.5 Å. For example, in HEWL-$(NAG)_3$-NAL the NAG in site B has no hydrogen-bonding interactions to HEWL. Although the pyranose ring does stack with the indole ring of Trp62, it does so from a mean distance of ~ 4.7 Å rather than the mean distance of 3.8 Å observed in the present work. Also in subsite D of the complex of $(NAG)_3$-NAL to HEWL, Asp52 makes very close contacts to the ring oxygen atom O5 and to the anomeric carbon atom, C1. In light of the very large differences observed between the refined and unrefined MGM complex structures, it is inappropriate to attach particular significance to the implications of the observed differences between MGM and $(NAG)_3$-NAL.

Environments of the Catalytic Residues

GLUTAMIC ACID 35. The conformation of the side chain of Glu35 is identical in the native HEWL structure and in the HEWL-MGM complex within the limits of accuracy of these determinations (Fig. 6.8). The side chain projects from the C-terminus of a helical segment of polypeptide that extends from Leu25 to Ser36 (Fig. 6.2, Fig. 6.15). In the native HEWL structure, the carboxyl group of Glu35 forms a hydrogen bond with the main chain NH group of Ala110 and another to a water molecule. In this latter hydrogen bond, Glu35 probably acts as a proton donor since neutron diffraction studies have confirmed the presence of a proton on the carboxyl group of Glu35 at the pH of the native crystals (4.2) (Mason *et al.*, 1984). In the HEWL-MGM complex, the carboxyl group also is the recipient of the hydrogen bond from the NH of Ala110 but now the other oxygen atom is hydrogen bonded to the anomeric oxygen (O1) of the reducing sugar in site D (Fig. 6.10c). Fig. 6.3, step b, indicates that the first event in the proposed catalytic pathway of HEWL is a transfer of a proton from a carboxyl oxygen of Glu35 to the glycosidic oxygen linking sugars D and E. Clearly, a preformed hydrogen bond between these two atoms (as described above) would facilitate this proton transfer event.

The carboxyl group of Glu35 is nestled into a relatively hydrophobic pocket formed by the side chains of Gln57 (the methylene carbon atoms $C\beta$, $C\gamma$), Trp108, Val109 and Ala110 (Fig. 6.10c, Fig. 6.15). In spite of this, it is the recipient of the hydrogen bond from Ala110 NH. In addition, the positive pole of a short helix (Val109 to Arg114) dipole points almost directly at the carboxyl group. This latter interaction could clearly stabilize a negative charge that develops on Glu35 during general acid catalysis (Fig. 6.3). Both the cluster of hydrophobic residues and the short helix are highly conserved in all the type C-lysozymes (Jollès & Jollès, 1984). Even in the T4 lysozyme (Weaver & Matthews, 1987), the catalytic glutamic acid, Glu11, has a relatively buried side chain in a hydrogen-bonded environment. In that enzyme however, the hydrogen bonding and negative charge stabilization come from a proximal arginine residue (Arg195) rather than a short helix dipole.

ASPARTIC ACID 52. The second amino-acid side chain that is strongly implicated in the catalytic hydrolysis of oligosaccharides is the carboxylate of Asp52. The side chain of Asp52 protrudes from the central strand of a three-stranded antiparallel β-sheet into the active site cleft (Fig. 6.16). In the native enzyme, one of the oxygen atoms ($O^{\delta 2}$) is solvent exposed and forms a hydrogen-bonded interaction to two solvent molecules that bridge across to the carboxyl group of Glu35. The other oxygen atom ($O^{\delta 1}$) is involved in a hydrogen-bonded network with the side chains of Asn46, Asp48, Ser50 and Asn59, all with their side chains on the same side of the β-sheet as the side chain of Asp52 (Fig. 6.16). This hydrogen-bonded network of five residues forms a "platform" against which the NAM ring and the carboxylate of the O3 lactyl group in site D pack (non-bonded contact distances ~ 3.4 to 4.0 Å).

Asp52 adopts a distinctly different conformation in the HEWL-MGM structure. In the complex, the "exposed" oxygen atom, $O^{\delta 2}$, has moved approximately 0.8 Å from its hydrated position in the native enzyme. It moves away from the carbohydrate ring of NAM in the D site. From this position it makes a hydrogen-bonded interaction with N2 of the acetamido group of ring D (3.2 Å) and a non-bonded contact to C1 of the same ring (3.3 Å). This change of conformation is achieved by rotations about χ^1 (7°) and χ^2 (-22°) that leave the other carboxylate oxygen atom ($O^{\delta 1}$) position unchanged and still part of the "platform" hydrogen-bonded network described for the native enzyme. Mimicking this type of motion it is possible to rotate χ^1 and χ^2 of Asp52 so that the exposed oxygen atom, $O^{\delta 2}$, comes to within 2.3 Å of C1 of the NAM residue in site D without moving $O^{\delta 1}$ from its hydrogen bonded position (χ_1 = -78°, χ_2 = -171°; see Fig. 6.16). Thus, it would seem to be possible to form a covalent bond between a nucleophilic Asp52 and the anomeric carbon of the NAM residue. The implications of this are discussed below.

The residues that comprise the hydrogen bonded network of the platform are all highly conserved in the sequences of C-type lysozymes (Jollès & Jollès, 1984). Thus, this network probably has a common function in the hydrolytic mechanism of these enzymes.

Implications of the HEWL-MGM Complex for the Catalytic Mechanism of Lysozyme

DISTORTION IN SITE D. An essential component of the mechanism whereby HEWL catalyzes the hydrolysis of oligosaccharides as put forward by Phillips and coworkers (Phillips, 1966, 1967; Blake *et al.*, 1967a,b; Imoto *et al.*, 1972) is the distortion of the NAM (or NAG) residue bound in site D (Fig. 6.3). Thus, the formation of the oxocarbonium transition state which, by definition requires C5, O5, C1 and C2 to be coplanar, is favored by the pretransition ground-state distortion of the sugar ring in site D into a twist-boat, $B_{3,0}$, or half-chain conformation. A variety of methods have been used to procure evidence for substrate distortion in the lysozyme mechanism.

Early on, free energies of association of a variety of sugars binding to the several subsites of HEWL were determined by fluorescence titration experiments (Chipman *et al.*, 1967, 1968, Chipman, 1971). Also, kinetic analyses of the cleavage rates of a variety of oligosaccharides were used to probe binding affinities of the subsites (Rupley *et al.*, 1967, Imoto *et al.*, 1972). These two studies concluded that the binding of sugar residues to site D was energetically unfavorable by approximately 3-6 kcal/mol. In both studies it was concluded that ground-state distortion of the saccharide in site D was the cause.

More direct structural data for distortion at site D has relied on the crystal structure of a complex between $(NAG)_3$-NAL and HEWL (Ford *et al.*, 1974). The δ-lactone of NAG was bound to site D and this transition state mimic provided strong evidence that site D favored sugar rings with sofa or $B_{3,0}$ conformations. Such conformations bring the C6-O6 hydroxymethyl group into a quasi-axial conformation (Fig. 6.11) and this was the observed conformation of the δ-lactone. Unfortunately, this structure determination was at a relatively low resolution (2.5 Å) and it has not been refined (see above).

A number of other studies do not support the idea of ground state distortion of the saccharide bound in site D. These studies are often cited as evidence that there is no distortion of the D ring (e.g. Post & Karplus, 1986). The initial report of the structure of the complex of HEWL-MGM (Kelly *et al.*, 1979) claimed that there was no distortion in

the D ring. Unfortunately that analysis was also based on an unrefined, relatively low-resolution (2.5 Å) crystal structure. We have shown with the present study that this initial interpretation was incorrect. The second structural study, often quoted as not supporting distortion of ring D, is an NMR analysis of the binding of MGM to HEWL in solution. Patt *et al.* (1978) did not observe any change in coupling constant between C1-H and C2-H of the reducing sugar NAM when it was bound to lysozyme. These authors measured the coupling constant between H_1^{α} H_2 and found that its value (2.6 hz) corresponded to the α-anomer (axial-OH) of NAM. Nonetheless, as discussed in a previous section, 3c(iii), and depicted in Fig. 6.13, the α-anomer of the NAM residue in site D cannot be accommodated in the present crystallographically determined NAM-binding mode due to the potentially extreme steric conflicts with Asp52.

Based on a detailed comparison of the binding constants of a series of oligosaccharide inhibitors of HEWL, Shindler and coworkers suggested that the O3-lactyl moiety on NAM is responsible for the majority of the unfavorable contribution to the binding free energy in subsite D (Shindler *et al.*, 1977). This would suggest that there is much less strain involved in binding a NAG residue to site D of HEWL. On the other hand, $(NAG)_3$-NAL which already has the residue to bind in site D in a distorted chair conformation (a sofa or half-chair), and $(NAG)_3$ NAX, which lacks the CH_2OH group thought to be responsible for the unfavorable contacts with lysozyme, are bound much more tightly than $(NAG)_3$ (Secemski & Lienhard, 1971; Secemski *et al.*, 1972; Van Eikeron & Chipman, 1972). Furthermore, the analysis of the kinetic parameters that resulted from the hydrolysis of p-nitrophenyl groups from a series of NAG and NAX containing oligosaccharides (Ballardie *et al.*, 1977), concluded that there must be an energetically favorable interaction between C6-O6 and the enzyme in the transition state to account for the observed behaviour of k_{cat}/k_m.

Theoretical chemical approaches have been used to evaluate the importance of strain in the hydrolysis of oligosaccharides by HEWL. Conformational energy calculations

performed on oligosaccharides binding in sites B, C and D of HEWL suggested that they could do so with no distortion of the ring in site D and that they can actually bind with a full $^4C_1$ chair conformation (Pincus &Sheraga, 1979, 1981). A similar conclusion was reached by Levitt from energy minimization procedures on a model of a hexasaccharide binding to all six subsites (A-F) of HEWL (Levitt, 1974). Molecular dynamics calculations on the lysozyme system have also been interpreted to rule out a role for "mechanical" distortion in the hydrolysis of substrates (Post & Karplus, 1986). The behaviour of lysozyme during a 55 psec simulation of a complex of $(NAG)_6$ bound in the active site lead to the proposal of an alternative pathway in which the endocyclic O5-C1 bond of ring D is broken initially. It is not clear from this proposed mechanism how the second substrate (H_2O) will be able to approach the anomeric carbon atom C1 in order to complete the hydrolysis.

It has been suggested that electrostatic fields are likely to play an important role in the hydrolysis of substrates (Warshel & Levitt, 1976; Warshel, 1981). More recently, it has been reported that the bond cleavage step in the enzymatic hydrolysis of oligosaccharides is promoted by a large electrostatic field across the active site cleft (Dao-Pin *et al.*, 1989). This field results from the charge distribution of the enzyme as a whole. The separation of charge in the general acid catalysis mechanism involves the transfer of a proton to the glycosidic oxygen from Glu35. If this step is favored by the electrostatic field across the active site cleft it is not easy to appreciate how the transfer of the proton from water back to Glu35 will be favorably influenced by this field.

The catalytic pathway proposed by Phillips and coworkers (Phillips, 1966, 1967; Blake *et al.*, 1967a,b) is summarized in Fig. 6.3a-e. The present crystal structure determination and refinement at 1.5 Å resolution of the complex between HEWL and MGM shows that ground state distortion of the carbohydrate ring bound in site D not only is possible, but that it does occur in this enzyme-product complex. The distortion is best described as a sofa conformation of the site D NAM. There are several interactions with

lysozyme that contribute to the distortion of the D ring. If one takes the conformation of NAM as observed in crystals of the α-anomer (Knox & Murthy, 1974) as the minimum energy conformation in solution, then disrupting the hydrogen-bonded interaction between the 2-acetamido-NH group and the carboxylate of the lactyl group on O3 is unfavorable. As the NAM residue approaches site D, the lactyl group is forced to move from its minimum energy hydrogen-bonded conformation to a position that places it between ring C and ring D due to the contacts with the "platform" (Asn46, Asp48, Ser50, Asp52, Asn59). Short contacts with O5 of NAG in site C make this a relatively unfavorable conformation but one that has a lower energy cost than the serious clash of the lactyl group in its α-anomeric disposition with residues on the "platform". Thus, our structure confirms a possible role of the lactyl group in an overall unfavorable binding energy in site D (Shindler *et al.*, 1977).

The well defined electron density and relatively low B-factors associated with the atoms of NAG in site C confirm the tight association that this saccharide has with HEWL. In the full $^4C_1$ chair conformation, the CH_2OH group on C5 of the NAM ring in site D makes what would be a much too close a contact with the relatively rigid acetamido group of ring C. We feel that this causes the observed distortion towards the quasi-axial position for C6-O6. A favorable hydrogen bond from the NH of Val109 to the O6 atom serves to help stabilize this conformation (Fig.6.10c). With NAM distorted in this way in the ground-state and the carboxyl group of Glu35 donating a proton to the glycosidic oxygen atom of the substrate, the groups are almost perfectly oriented for protonation of the glycosidic oxygen atom resulting in the cleavage of the C1-O4' bond to form the oxocarbonium ion intermediate (Fig. 6.3c). Small shifts of the atoms of the NAM ring place C5, O5, C1 and C2 in a plane with the C6-O6 group quasi-axial, a conformation that was imposed by the binding to site D. Since the potential barrier for such a ring distortion is approximately 6-10 kcal/mol (Anet & Brown, 1967) the enzyme has a profound role in catalysis by inducing the conformational change in the substrate. The oxocarbonium ion

intermediate is stabilized by the proximity of the formal negative charges on Glu35 and Asp52. In turn, the negative charge on the Glu35 carboxylate is stabilized by the hydrogen-bond from the NH group of Ala110 and by the fact that it is positioned at the positive pole of a helix dipole (Val109-Arg114). The negative charge on Asp52 is stabilized by hydrogen-bonding interactions from the amide side chains of Asn46 and Asn59.

A water molecule would replace the leaving hydroxyl group HO4 of ring E and be in an appropriate position for OH^- attack on C1 of ring D upon depositing its proton with the carboxylate of Glu35 (Fig. 6.3d). This direction of attack by the OH^- ion results in retention of the equatorial β-anomer of the resulting NAM-OH in site D.

IS THERE A COVALENT GLYCOSYL-ENZYME INTERMEDIATE? There has been much written on the mechanism of HEWL and whether or not an oxocarbonium ion intermediate stabilized by the proximity of the negatively charged carboxylate of Asp52 is feasible (Sinnott, 1987). The mechanism involving covalent catalysis through a glycosyl-enzyme intermediate was first elaborated by Koshland (1953). All of the necessary elements for hydrolysis via this pathway seem to be in place in the lysozyme system. In addition, there are several (e→e) β-glycosidases that almost certainly catalyze the hydrolysis of substrates through a covalently attached intermediate (Sinnott, 1987).

We have explored the possibility of forming a covalently-attached glycosyl enzyme through the formation of a bond from Asp $COO^{\delta 2}$ (the nucleophile) and C1 of the oxocarbonium ion intermediate. As we have pointed out in section e(ii), it is possible to adjust χ^1 and χ^2 of Asp52 to allow a close approach of 2.3 Å between $O^{\delta 2}$ of Asp52 and the anomeric carbon atom of the NAM residue in site D (Fig. 6.16a). However, an additional ~ 0.8 Å movement is required before a covalent bond can be formed. This movement is analogous to that required in the formation of the covalent tetrahedral intermediate on the catalytic pathway of the serine proteinases (James *et al.*, 1980). In that pathway, the nucleophilic alkoxide ion CH_2-O^- of Ser195 moves into position as the

carbonyl carbon of the scissile bond approaches and takes on tetrahedral character. The resultant covalent C-O bond is of the order of 1.5 Å in length.

Fig. 6.16a shows the possible direction of approach of $O^{\delta 2}$ of Asp52 to the anomeric carbon atom C1. Asp52 $O^{\delta 1}$ is relatively fixed in its hydrogen-bonded position to the platform. Thus, unless the hydrogen-bonded interactions to Asn46 and Asn59 that maintain this position are disrupted, the carboxyl group cannot approach more closely than 2.3 Å to the C1 atom of NAM. It is not likely that the NAM residue will move much more towards Asp52 either, as too-short non-bonded contacts would then occur between the hydrophobic side of the ring (C3-H and C5-H) and the residues of the platform (Fig. 6.10c). Thus, the distance of closest approach appears to be ~ 2.0 Å.

A second argument against forming a covalently attached glycosyl intermediate comes from the direction of approach of $O^{\delta 2}$ of Asp52 in forming the hypothetical acylal intermediate. The formation of esters of carboxylic acids takes place almost exclusively from the *syn* direction rather than the *anti* as a result of the stronger basicity of the carboxylate nucleophile in this direction (Gandour, 1978). It can be surmised from an examination of Fig. 6.16a that the *syn* lone pair orbitals on Asp52 carboxylate are not favorably oriented for forming a covalent acylal bond to C1 of the D sugar. The direction of approach of the carboxylate oxygen atom $O^{\delta 2}$ (i.e. C^{γ}-$O^{\delta 2}$···C1) is 170° rather than the most favorable angle of 120°. In addition, the most favorable conformation for the acylal adduct is that shown in Fig. 6.16b. It would not be possible to achieve such a conformation without completely disrupting the observed interactions formed by Asp52 and dramatically changing the conformation of the supporting strands in the β-sheet.

Although present techniques in X-ray crystallography cannot give information about lifetimes of intermediates or about how rapidly enzyme catalyzed reactions take place, the results of this most powerful technique provide details of the relative disposition of atoms which must be taken into account when contemplating various reaction mechanisms. The results of the present crystallographic analysis of the complex between HEWL and the

trisaccharide MGM are most consistent with the general acid hydrolysis mechanism originally suggested by Phillips and coworkers. It also shows how important it is to determine crystal structures of enzymes and their complexes at as high a resolution as possible and to refine these structures to the limit of the available diffraction data.

ACKNOWLEDGEMENTS

Dr. Brian Sykes kindly provided the sample of NAM-NAG-NAM. We wish to thank Stanley Moore for growing the crystals used in this study and for assistance in data collection. Dr. Anita Sielecki provided advice during refinement stages. Mae Wylie expertly and patiently dealt with many revisions of the text. This work was supported in part by the Medical Research Council of Canada. N.C.J.S. is a recipient of an Alberta Heritage Foundation for Medical Research Studentship.

REFERENCES

Anet, F.A.L. & Brown, J.R. (1967). *J. Am. Chem. Soc.* **89**, 760-765.

Artymiuk, P.J., Blake, C.C.F., Rice, D.W. & Wilson, K.S. (1982). *Acta Cryst. B* **38**, 778-783.

Ballardie, F.W., Capon, B., Cuthbert, M.W. & Dearie, W.M. (1977). *Bioorganic Chem.* **6**, 483-509.

Banjeree, S.K., Holler, E., Hess, G.P. & Rupley, J.A. (1975). *J. Biol. Chem.* **250**, 4355-4367.

Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F. Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535-542.

Blake, C.C.F., Koenig, D.F., Mair, G.A., North, A.C.T., Phillips, D.C. & Sarma, V.R. (1965). *Nature* **206**, 757-761.

Blake, C.C.F., Mair, G.A., North, A.C.T., Phillips, D.C. & Sarma, V.R. (1967a). *Proc. Roy. Soc. London B* **167**, 365-377.

Blake, C.C.F., Johnson, L.N., Mair, G.A., North, A.C.T., Phillips, D.C. & Sarma, V.R. (1967b). *Proc. Roy. Soc. London B* **167**, 378-388.

Cambillau, C.M. (1987). "Documentation for TOM, Molecular Graphics Program for the IRIS. CRMC2, Campus Luminy, Case 913, Marseille, France.

Canfield, R.E. (1963). *J. Biol. Chem.* **238**, 2698-2707.

Cheetham, J.C., Artymiuk, P.J. & Phillips, D.C. (1990). *J. Mol. Biol.* (in press).

Chipman, D.M. (1971). *Biochemistry* **10**, 1715-1722.

Chipman, D.M. & Sharon, N. (1969). *Science* **167**, 454-469.

Crawford, J.L., Lipscomb, W.N. & Schellman, C.G. (1973) *Proc. Natl. Acad. Sci.* **70**, 538-592.

Cruickshank, D.W.J. (1949). *Acta Cryst.* **2**, 65-82.

Cruickshank, D.W.J. (1954). *Acta Cryst.* **7**, 519-523.

Cruickshank, D.W.J. (1967). In *International Tables for X-ray Crystallography* (Kasper, J.S. & Lonsdale, K., eds.), Vol. II, pp. 318-340, Kynoch Press, Birmingham, England.

Dahlquist, F.W., Borders, C.L., Jacobson, G. & Raftery, M.A. (1969a) *Biochemistry* **8**, 694-699.

Dahlquist, F.W., Rand-Meir, T. & Raftery, M.A. (1969b). *Biochemistry* **8**, 4214-4221.

Dao-Pin, S., Liao, D., Remington, S. (1989). *Proc. Natl. Acad. Sci.* **86**, 5361-5365.

Ford, L.O., Johnson, L.N., Machin, P.A., Phillips, D.C. & Tjian, R. (1974). *J. Mol. Biol.* **88**, 349-371.

Fujinaga, M., Delbaere, L.T.J., Brayer, G.D. & James, M.N.G. (1985). *J. Mol. Biol.* **183**, 479-502.

Gandour, R.D. (1978). In *Transition States of Biochemical Processes* (Gandour, R.D. & Schowen, R.L., eds.), pp. 529-552, Plenum Press, New York.

Hamlin, R. (1985). *Methods Enzymol.* **114**, 416-452.

Hendrickson, W.A. & Konnert, J.H. (1980). In *Computing in Crystallography* (Diamond, R., Ramaeshan, S., Venkatesan, K., eds.), pp. 13.01-13.23, Indian Academy of Sciences, International Union of Crystallography, Bangalore, India.

Hodsdon, J.M., Brown, G.M., Sieker, L.C. & Jensen, L.H. (1990). *Acta Cryst.* **B46**, 54-62.

Hogle, J., Rao, S.T., Mallikarjunan, M., Beddell, C., McMullan, R.K. & Sundralingam, M. (1981). *Acta Cryst. B* **37**, 591-597.

Holler, E., Rupley, J.A. & Hess, G.P. (1975a). *Biochemistry* **14**, 1088-1094.

Holler, E., Rupley, J.A. & Hess, G.P. (1975b). *Biochemistry* **14**, 2377-2385.

Howard, A.J., Nielson, C. & Xuong, N.H. (1985). *Methods Enzymol.* **114**, 452-472.

Imoto, I., Johnson, L.N., North, A.C.T., Phillips, D.C. & Rupley, J.A. (1972). In *The Enzymes* (Boyer, P., ed.), Vol. VII, 3rd edn., pp. 665-868, Academic Press, New York.

Inaka, K., Matsushima, M. & Morikana, K. (1990). Collected Abstracts of the XV Congress and General Assembly of the IUCr. PS-03.03.20.

James, M.N.G., Sielecki, A.R. & Moult, J. (1983). In *"Peptides, Structure and Function"*, (Hruby, V. & Rich, D., eds.), Pierce Chemical Co., Rockford, Ill., pp. 521-530.

James, M.N.G. & Sielecki, A.R. (1987). In *Biological Macromolecules and Assemblies, Active Sites of Enzymes*, Vol. 3, (Jurnak, F.A. & McPherson, A., eds.), John Wiley & Sons, New York, pp. 413-482.

James, M.N.G., Sielecki, A.R., Brayer, G.D., Delbaere, L.T.J. & Bauer, C.-A. (1980). *J. Mol. Biol.* **144**, 43-89.

Jeanloz, R.W., Sharon, N. & Flowers, H.M. (1963). *Biochim. Biophys. Res. Comm.* **13**, 20-25.

Jefferey, G.A. & Taylor, R. (1980). *J. Comput. Chem.* 1, 99-109.

Jefferey, G.A. (1990). *Acta Cryst.* **B46**, 89-103.

Johnston, L.N., Cheetham, J., McLaughlin, P.J., Acharya, K.P., Barford, D. & Phillips, D.C. (1988). *Current Topics in Microbiology and Immunology* **139**, 81-

129.

Jollès, J., Jauregui-Adell, J., Bernier, I. & Jollès, P. (1963). *Biochim. Biophys. Acta* **78**, 668-689.

Jollès, J. & Jollès, P. (1961). *CR Acad. Sci. Paris* **253**, 2773-2775.

Jollès, P. & Jollès, J. (1984). *Molecular and Cellular Bioch.* **63**, 165-189.

Jones, T.A. (1985). *Meth. Enzym.* **115**, 157-171.

Joynson, M.A., North, A.C.T., Sarma, V., Dickerson, R.E. & Steinrouf, L.K. (1970). *J. Mol. Biol.* **50**, 137-142.

Kelly, J.A., Sielecki, A.R., Sykes, B.D., James, M.N.G. & Phillips, D.C. (1979). *Nature* **282**, 875-878.

Knox, J.R. & Murthy, N.S. (1974). *Acta Cryst.* **B30**, 365-371.

Koshland, D.E. (1953). *Biol. Rev.* **28**, 416-420.

Kravachenko, N.A. (1967). *Proc. Roy. Soc. London* **B167**, 429-430.

Kundrot, C.E. & Richards, F.M. (1987). *J. Mol. Biol.* **193**, 157-170.

Kuramitsu, S., Ikeda, K. & Hamaguchi, K. (1977). *J. Biochem.* **82**, 585-597.

Kuroki, R., Yamada, H., Moriyama, T. & Imoto, T. (1986). *J. Biol. Chem.* **261**, 13571-13574.

Levitt, M. (1974). In *Peptides, Polypeptides and Proteins* (Blout, E.R., Bovey, F.A., Goodman, M. & Lotan, N., eds.), pp. 99-113, Wiley & Sons, New York.

Levitt, M., Sander, C. & Stern, P.S. (1985). *J. Mol. Biol.* **181**, 423-447.

Lin, T. & Koshland, D.E. (1969). *J. Biol. Chem.* **244**, 505-508.

Lin, T.-Y. (1970). *Biochemistry* **9**, 984-994.

Malcolm, B.A., Rosenberg, S., Corey, M.J., Allen, J.S., de Baetselier, A. & Kirsch, J.F. (1989). *Proc. Natl. Acad. Sci.* **86**, 133-137.

Mason, S.A., Bentley, G.A. & McIntyre, G.J. (1984). In *Neutrons in Biology* (Shoenborn, B., ed.), pp. 323-334, Plenum Press, New York.

Moult, J., Yonath, A., Traub, W., Smilansky, A., Podjarny, A., Rabinovich, D. & Saya, A. (1976). *J. Mol. Biol.* **100**, 179-195.

Mo, F. & Jensen, L.H. (1978). *Acta Cryst.* **B34**, 1562-1569.

Nguyen-Huu, M.C., Stratmann, M., Groner, B., Wurtz, T., Land, H., Glesecke, K., Sippel, A.E. & Schutz, G. (1979). *Proc. Natl. Acad. Sci.* **76**, 76-80.

Pahler, A. & Hendrickson, W.A. (1990). *J. Appl. Cryst.* **23**, 218-221.

Parsons, S.M. & Raffery, M.A. (1969). *Biochemistry* **8**, 4199-4205.

Parsons, S.M. & Raffery, M.A. (1970) *Biochim. Biophys. Res. Comm.* **41**, 45-49.

Patt, S.L., Baldo, J.H., Boekelheide, K., Weisz, G. & Sykes, B.D. (1978). *Can. J. Biochem.* **56**, 624-629.

Perkins, S.J., Johnson, L.N., Machin, P.A. & Phillips, D.C. (1978). *Biochem. J.* **173**, 607-616.

Phillips, D.C. (1966). *Scientific American* **215**, 78-90.

Phillips, D.C. (1967). *Proc. Roy. Acad. Sci.* **57**, 484-495.

Pincus, M.R. & Sheraga, H.A. (1979). *Macromolecules* **12**, 633-644.

Pincus, M.R. & Sheraga, H.A. (1981). *Biochemistry* **20**, 3960-3965.

Post, C.B. & Karplus, M. (1986). *J. Amer. Chem. Soc.* **108**, 1317-1319.

Post, C.B., Brooks, B.R., Karplus, M., Dobson, C.M., Artymiuk, P.J., Cheetham, J.C. & Phillips, D.C. (1986). *J. Mol. Biol.* **190**, 455-479.

Post, C.B., Dobson, C.M. & Karplus, M. (1989). *Proteins: Structure, Function and Genetics* **5**, 337-354.

Quiocho, F.A. (1988). *Curr. Topics in Microbiol. and Immunol.* **126**, 135-148.

Quiocho, F.A. (1986). *Ann. Rev. Biochem.* **55**, 287-315.

Quiocho, F.A. (1989). *Pure and Applied Chem.* **61**, 1293-1306.

Ramakrishnan, C. & Ramachandran, G.N. (1965). *Biophys. J.* **5**, 909-933.

Read, R.J., Fujinaga, M., Sielecki, A.R. & James, M.N.G. (1983). *Biochemistry* **22**, 4420-4433.

Rosenberg, S. & Kirsch, J.F. (1981). *Biochemistry* **20**, 3196-3204.

Rupley, J.A. (1967). *Proc. Roy. Soc. Sect. B* **167**, 416-428.

Rupley, J.A., Butler, L., Gerring, M., Hartdegen, F.J. & Pacoravo, R. (1967) *Proc. Natl. Acad. Sci.* **57**, 1088-1095.

Salton, M.R.J. & Ghuysen, J.M. (1959). *Biochim. Biophys. Acta* **36**, 552-554.

Salton, M.R.J. & Ghuysen, J.M. (1960). *Biochim. Biophys. Acta* **45**, 355-363.

Secemski, I.I. & Leinhard, G.E. (1971). *J. Amer. Chem. Soc.* **93**, 3549-3550.

Secemski, I.I., Lehrev, S.S. & Lienhard, G.E. (1972). *J. Biol. Chem.* **247**, 4740-4748.

Shindler, M., Assaf, Y., Sharon, N. & Chipman, D.M. (1977). *Biochemistry* **16**, 423-431.

Sinnott, M.L. (1987). In *Enzyme Mechanisms* (Page, M.I. & Williams, A., ed.), pp. 259-291, Royal Soc. Chem., London.

Smith, L.E.H., Mohr, L.H. & Raftery, M.A. (1973) *J. Amer. Chem. Soc.* **95**, 7497-7501.

Van Eikeron, P. & Chipman, D.M. (1972). *J. Am. Chem. Soc.* **94**, 4788-4790.

Warshel, A. & Levitt, M. (1976). *J. Mol. Biol.* **103**, 227-249.

Warshel, A. (1981) *Acc. Chem. Res.* **14**, 284-291.

Weaver, L.H. & Matthews, B.W. (1987). *J. Mol. Biol.* **193**, 189-198.

Wilson, A.J.C. (1942). *Nature* **150**, 151-152.

Xuong, N.H., Sullivan, D., Nielson, C., Hamlin, R. & Anderson, D. (1985). *J. Appl. Cryst.* **18**, 342-350.

Table 1

Summary of observed intensities for the lysozyme-MGM complex

Resolution range	No. of possible reflections	No. of measured reflections	No. of reflections ($I \geq 2\sigma_I$) (% total)
∞ - 2.5	4285	4280	97%
2.5 - 2.0	4077	4069	96%
2.0 - 1.77	4049	4023	92%
1.77 - 1.60	3995	3960	87%
1.60 - 1.49	3963	2874	57%

Table 2

Refinement Statistics

	Native	Complex
Resolution range (Å)	8.0-1.75	8.0-1.5
R (= $\Sigma \|F_o\| - \|F_c\| / \Sigma\|F_o\|$)	0.154	0.169
No. of reflections ($\|F\| \geq 4\ \sigma_F$)	10315	17386
R (all data)	18.9	19.1
No. of protein atoms	1000	1000
No. of sugar atoms	-	53
No. of solvent atoms	127	149
r.m.s. deviations from ideal values		
Root mean square error		
Bond distances (Å)	.017	.023
Angle distances (Å)	.037	.043
Planar 1-4 distances (Å)	.040	.046
Planar groups (Å)	.010	.017
Chiral centers (Å^3)	.162	.179
Non-bonded contacts		
(a) single torsions (Å)	.235	.241
(b) multiple torsions (Å)	.190	.232
(c) hydrogen bonds (Å)	.236	.238

Table 3

Average temperature factors for saccharides bound to sites B, C and D of HEWL

Sugar Moiety	B <B>† (Å^2)	C <B>† (Å^2)	D <B>† (Å^2)
Ring atoms (C1,C2,C3,C4,C5,O5)	36	22	39
Exocyclic oxygen atoms			
O1	-	-	36
O3	49	17	40
O4	41	29	28
Hydroxymethyl atoms			
C6	35	29	40
O6	36	32	29
Acetamido atoms (N2,C7,O7,C8)	45	15	58
Lactyl atoms (C9,C10,C11,O10,O12)	56	-	47

† Average temperature factor values (<B>) are given when more than one atom is included in the group.

Table 4

Geometry of glycosidic linkages*

Sugars	ϕ (°)	ψ (°)	ψ_H (°)	O5·H-O3' (Å)
MGM				
B-C	-89	-120	30	3.0
C-D	-108	-110	11	-
GGG†				
A-B	-40	-126	70	3.5
B-C	-90	-117	28	3.2

* Values of ϕ,ψ are the torsional angles about C1-O4' and O4'-C4' defined by O5-C1-O4'-C4' and C1-O4'-C4'-C5', respectively. The value of ψ_H is the helical twist parameter defined as the average of the pseudorotation angles ψ_1 = O5-C1-C4'-C3' and ψ_2 = C2-C1-C4'-C5' (Mo & Jensen, 1978).

† These values are from the structure of $(NAG)_3$ binding into sites A, B and C of HEWL (Cheetham, Artymiuk & Phillips, 1990).

Table 5

Summary of hydrogen bonded contacts in the HEWL-MGM complex

	Direct hydrogen bond to protein			Water mediated hydrogen bond to protein		
Sugar ring	Atom	Atom	Dist (Å)	Atom	Atom	Dist (Å)
B	O6	Asp101 OD1	2.4	O6	O236	2.6
	O10	Asn103 ND2	2.8	O236	O163	2.8
	O12	Asn103 ND2	3.2	O163	Gly104 O	2.9
				O163	Ile98 O	2.9
				O7	O270	3.2
				O270	O236	2.8
C	O3	Trp63 NE1	3.0	O6	O205	3.1
	O6	Trp62 NE1	2.9	O205	Asp48 OD2	2.9
	N2	Ala107 O	2.9			
	O7	Asn59 N	2.9			
D	O1	Glu35 OE1	2.9	O1	O181	2.7
	O6	Val109 N	2.8	O181	Asn57 OE1	2.8
	N2	Asp52 OD2	3.2	O10	O200	2.3
	N2	Asn46 OD1	3.4	O200	Asp48 OD1	3.1
				O200	Thr47 OG1	3.2
				O10	O205	2.4
				O205	Asp48 OD2	2.9

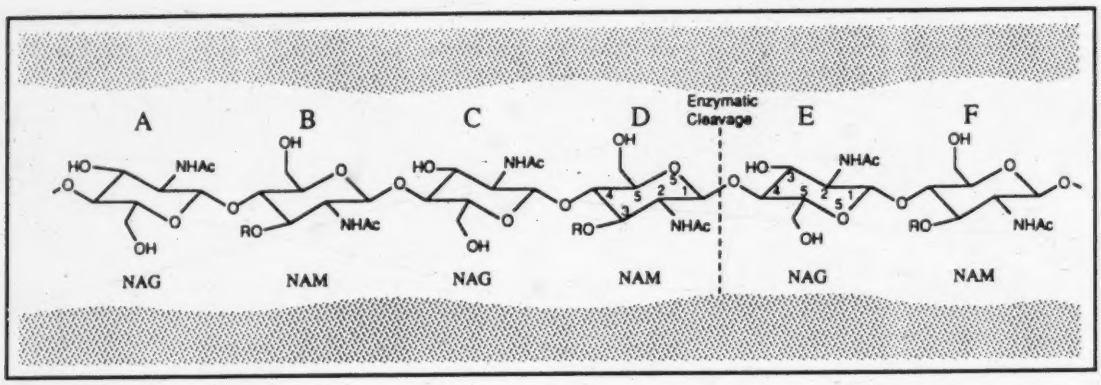


Fig. 6.1 The hexasaccharide comprised of alternating 2-acetamido-2-deoxy-D-glucose (NAG) and 2-acetamido-2-deoxy-D-muramic acid (NAM) residues as found in the natural cell wall peptidoglycan of gram negative bacteria. NHAc refers to the N-acetyl or acetamido moiety found in both NAG and NAM sugars. R refers to the lactyl or $OCH(CH_3)COOH$ moiety found only in NAM sugars. The six saccharide units are thought to bind to successive sites on lysozyme labeled A through F.

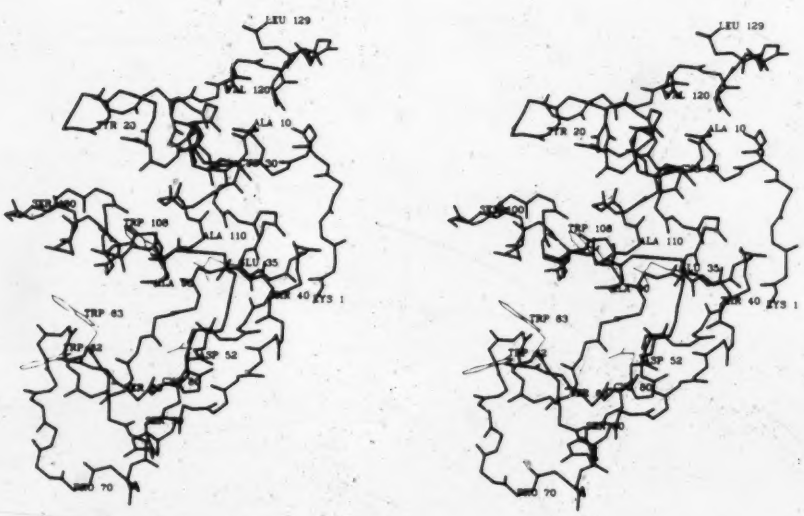
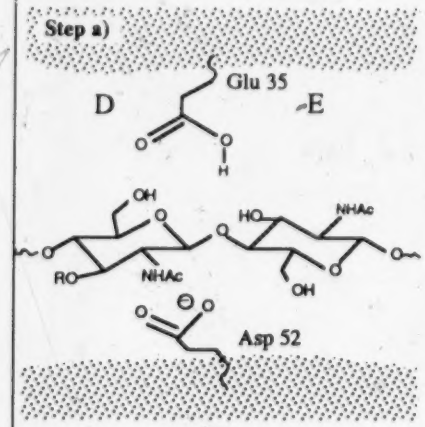
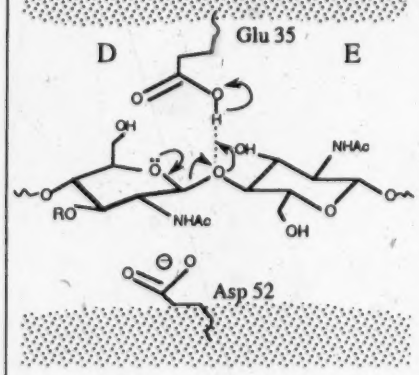


Fig. 6.2 A stereographic representation highlighting the overall fold of hen egg white lysozyme. For clarity only the atoms of main chain and selected side chains within the active site region are shown.

Figure 3

Figure 3

Proposed Mechanism	Supporting Evidence	Reference
Step a)  The two acidic residues, Glu35 and Asp52, flank the glycosidic oxygen atom that links the NAM residue at site D and the NAG residue at site E. The specificity should be for the cleavage between the D and E sugars. Due to the hydrophobic nature of its environment in HEWL, Glu35 would be protonated at the pH of maximal enzymatic activity (pH5). Asp52 would be ionized as it is in a hydrophilic environment.	- All sequences of type-C lysozymes have a Glu and Asp in analogous positions to Glu35 and Asp52. - Chemical modification of Asp52 leads to inactivation. - Site specific mutagenesis of Glu35 and Asp52 totally inactivates lysozymes ability to cleave the natural hexasaccharide substrate. - The specificity of enzymic hydrolysis is for the bond between the fourth and fifth sugar units from the non-reducing end of the natural substrate. - pH difference spectra of native HEWL and derivatives with either Glu35 or Asp52 esterified indicate pKa values of 6.1-6.5 for Glu35 and 4.7-5.1 for Asp52. - Neutron diffraction to 1.4 Å resolution shows a peak of negative density (i.e. a hydrogen) off of the carboxylate oxygen of Glu35.	*Jollès & Jollès, 1984* *Parsons & Raftery, 1969; Lin et al., 1969, 1970 Kuroki et al., 1986* *Malcolm et al., 1987* *Imoto et al., 1972; Holler, et al., 1975a; Banjeree et al., 1975* *Parsons & Raftery, 1970; Kuramitsu et al., 1977* *Mason et al., 1984*
Step b)  An optimal fit of the hexasaccharide substrate into the active site cleft requires a steric distortion of the sugar ring at site D into the higher energy half chair conformation thereby weakening the bond between atoms C1 and O4 of residues D and E (i.e. creating a higher energy ground state). Glu35 acting as a general acid catalyst would donate a proton to the weakened glycosidic bond between sugar residues D and E.	- The crystallographic determination of a lactone derivative of tetra (NAG) bound to sites A to D of lysozyme indicates the most probable conformation of site D was a sofa or boat. Not possible to distinguish between the two due to low resolution and lack of refinement. - A positive free energy term was associated with the binding of a NAM saccharide at site D.	*Ford et al., 1974* *Chipman & Sharon, 1969; Chipman, 1971; Rupley, 1967; Imoto et al., 1972*

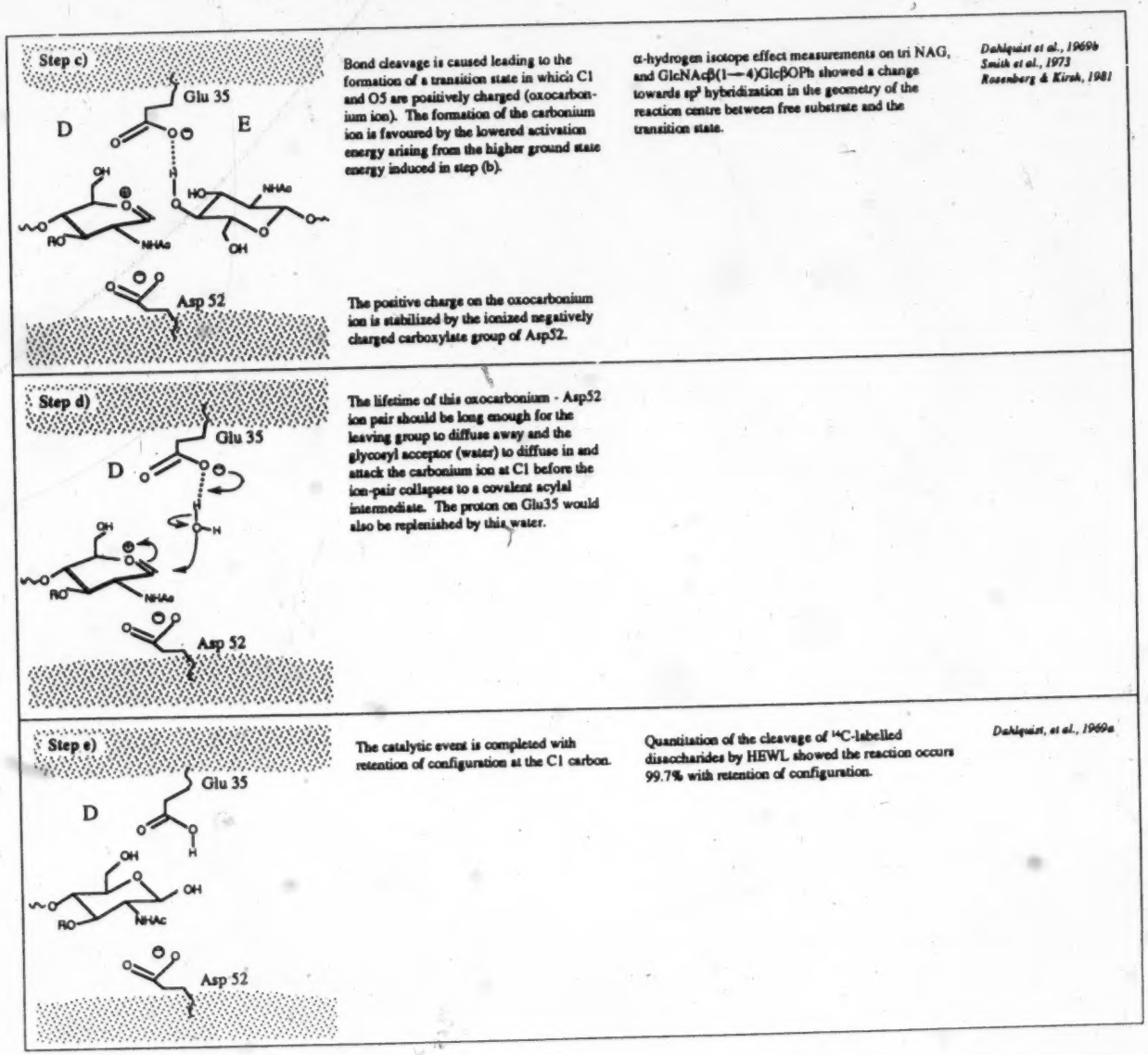

Step c)
Glu 35
D
E
Asp 52
Bond cleavage is caused leading to the formation of a transition state in which C1 and O5 are positively charged (oxocarbonium ion). The formation of the carbonium ion is favoured by the lowered activation energy arising from the higher ground state energy induced in step (b).
The positive charge on the oxocarbonium ion is stabilized by the ionized negatively charged carboxylate group of Asp52.
α-hydrogen isotope effect measurements on tri NAG, and GlcNAcβ(1—4)GlcβOPh showed a change towards sp³ hybridization in the geometry of the reaction centre between free substrate and the transition state.
Dahlquist et al., 1969b
Smith et al., 1973
Rosenberg & Kirsh, 1981
Step d)
Glu 35
D
Asp 52
The lifetime of this oxocarbonium - Asp52 ion pair should be long enough for the leaving group to diffuse away and the glycosyl acceptor (water) to diffuse in and attack the carbonium ion at C1 before the ion-pair collapses to a covalent acylal intermediate. The proton on Glu35 would also be replenished by this water.
Step e)
Glu 35
D
Asp 52
The catalytic event is completed with retention of configuration at the C1 carbon.
Quantitation of the cleavage of ¹⁴C-labelled disaccharides by HEWL showed the reaction occurs 99.7% with retention of configuration.
Dahlquist, et al., 1969a

Fig. 6.3 The general acid catalytic pathway of HEWL as proposed by Phillips (1966).

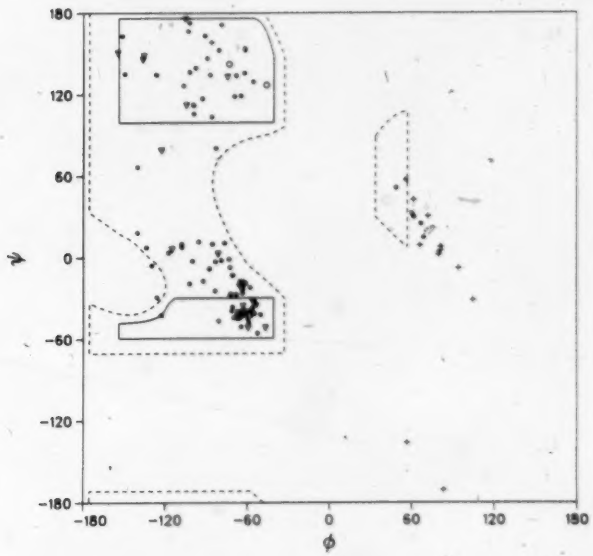


Fig. 6.4 A plot of ϕ vs ψ [Ramakrishnan & Ramachandran (1965)] for the refined trisaccharide-HEWL complex structure. Gly residues are denoted by (+); Pro by (o); β-branched residues by (∇) and all others by (•).

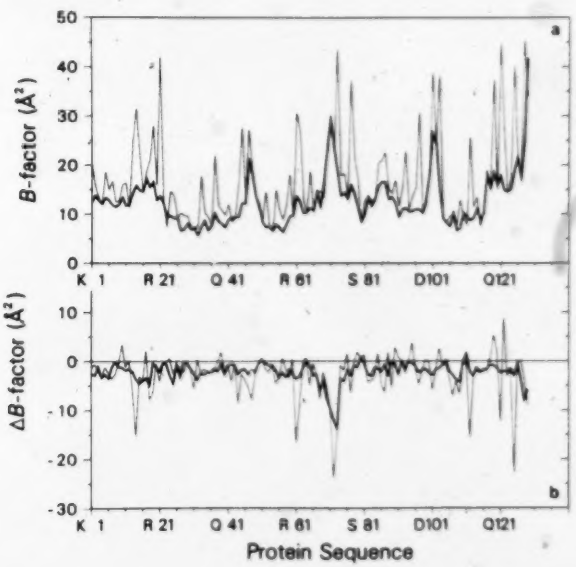
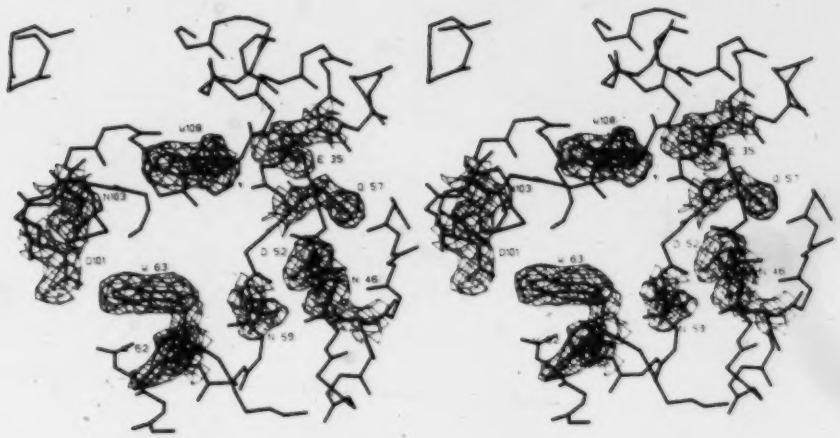


Fig. 6.5 (a) A plot of the mean temperature factors for the 1000 protein atoms of the lysozyme molecule in the trisaccharide complex form. Main chain atoms are indicated in thick lines and side chain atoms in thin lines. (b) A plot of the differences in mean temperature factors for the saccharide complexed and native lysozyme forms. The ΔB value plotted at each residue position corresponds to the B value of the complex minus the B value of the native.

a



b

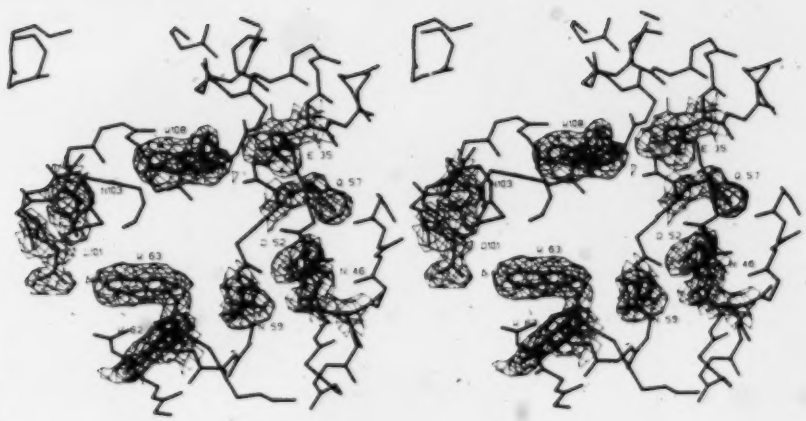
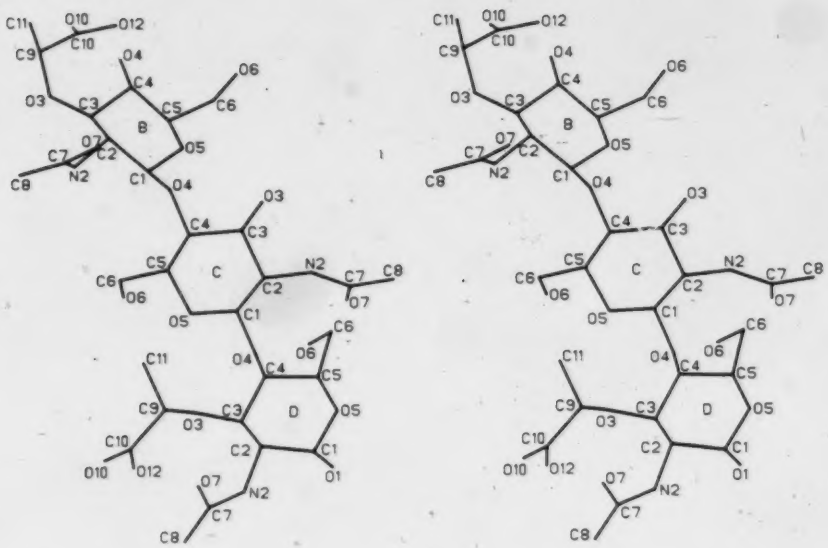


Fig. 6.6 The active site region of HEWL in the (a) native and (b) MGM complexed forms. For clarity only the electron density of select side chains is shown. The electron density has been calculated with coefficients $2|F_o|-|F_c|$ and the phases of the final least squares cycle.. The map is contoured at +0.40 eÅ^{-3}.

a



b

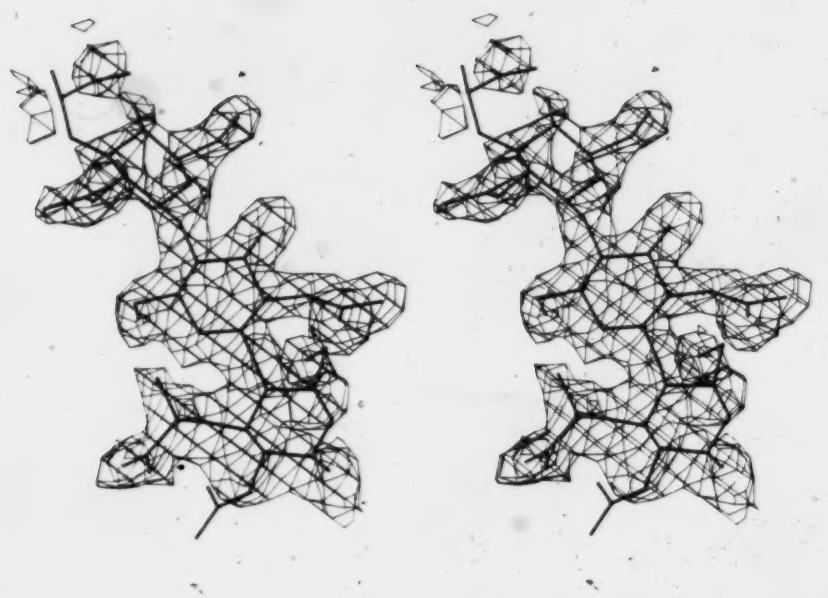


Fig. 6.7 (a) The final, refined enzyme bound conformation of the trisaccharide NAM-NAG-NAM. All atoms within sugar units B, C, and D are labelled appropriately for reference in the text. (b) The electron density map computed in the region of the trisaccharide using coefficients $2|F_o|-|F_c|$ and phases α_c from the final least-squares cycle. The map is contoured at +0.30 eÅ^{-3}.

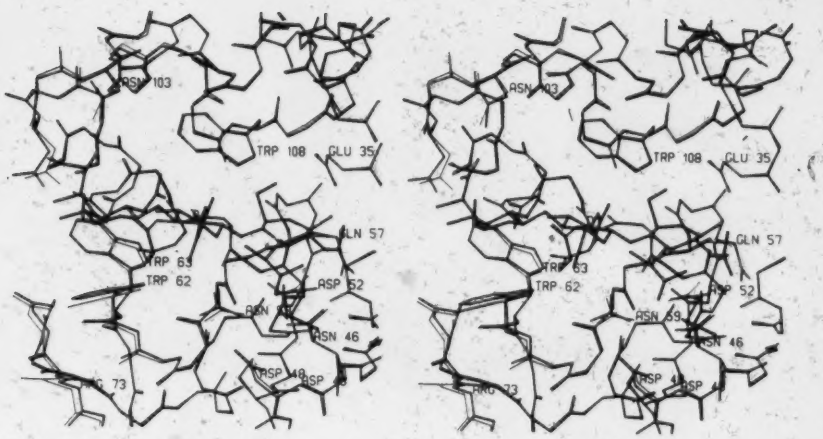


Fig. 6.8 An overlap of the refined native lysozyme (thin lines) and sugar bound lysozyme (thick lines), emphasizing the very subtle conformational changes that the enzyme molecule undergoes to accommodate the trisaccharide. The result is a narrowing of the cleft by ~ 1.0 Å overall.

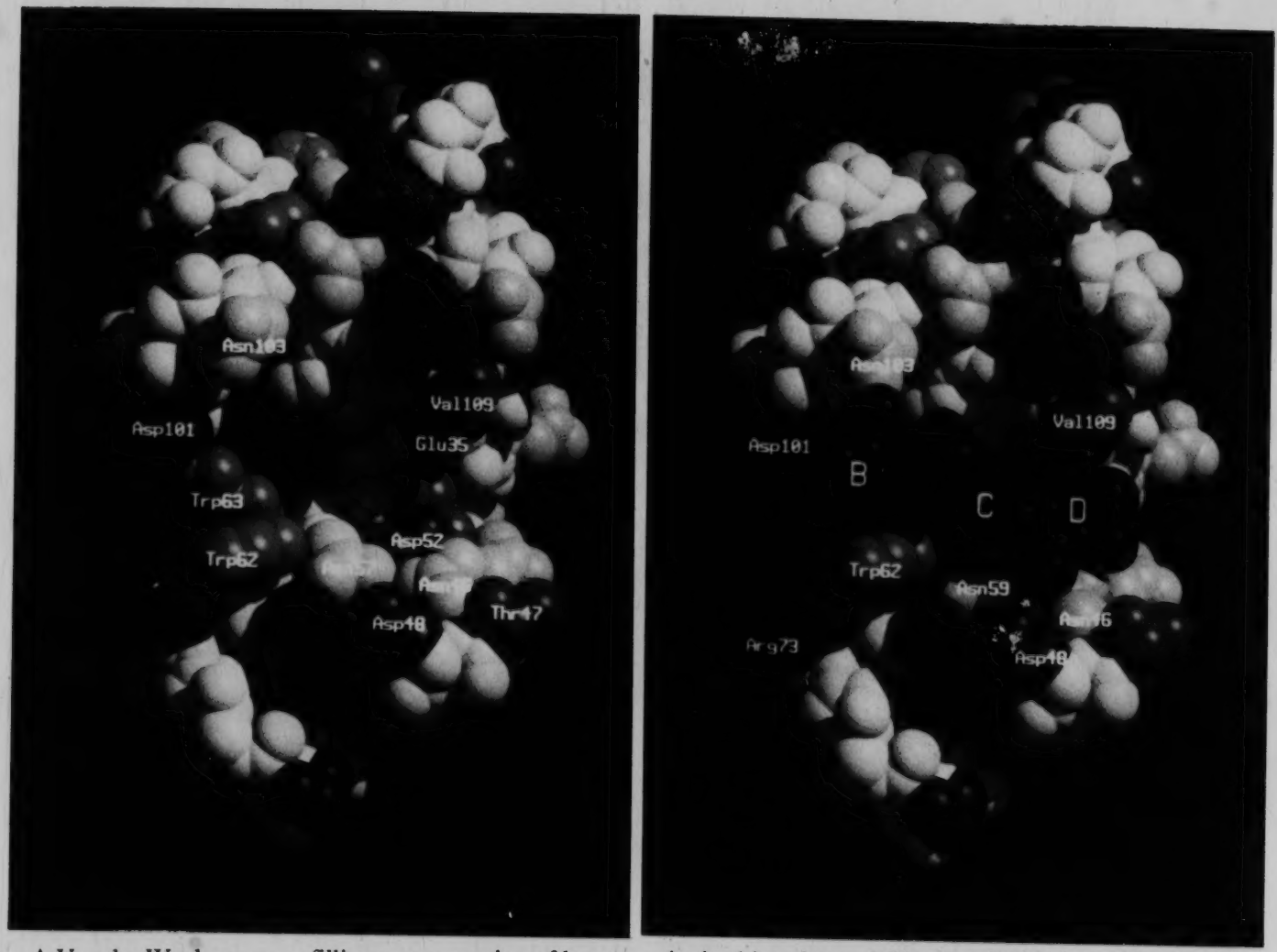
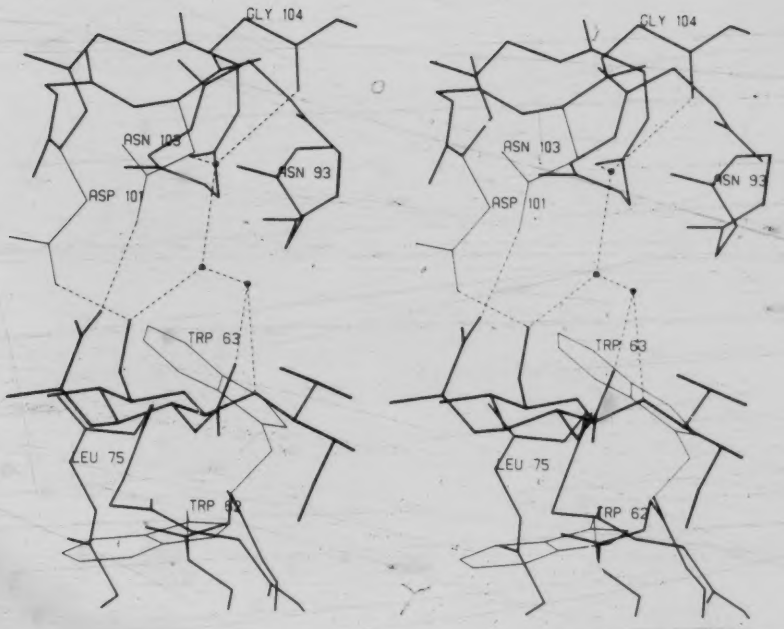
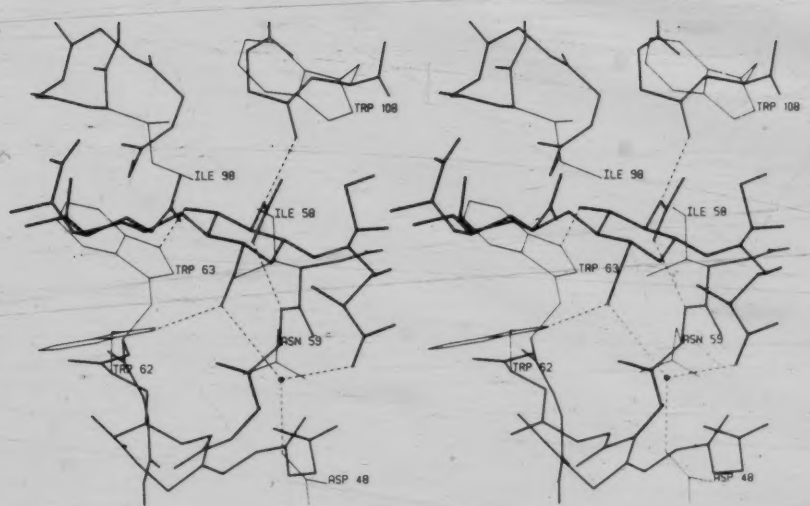


Fig. 6.9 A Van der Waals or space filling representation of lysozyme in the (a) native and (b) trisaccharide-bound forms. The view is looking directly into the active site cleft (similar to that shown in Fig. 6.8). Main chain atoms are in white, acidic side chain atoms in red, basic side chain atoms in blue, hydrophobic, aliphatic side chains in green, aromatic side chains in brown. The trisaccharide in (b) is purple.


GLY 104
ASN 103
ASN 93
ASP 101
TRP 63
LEU 75
TRP 62
GLY 104
ASN 103
ASN 93
ASP 101
TRP 63
LEU 75
TRP 62

a


TRP 108
ILE 98
ILE 58
TRP 63
ASN 59
TRP 62
ASP 48
TRP 108
ILE 98
ILE 58
TRP 63
ASN 59
TRP 62
ASP 48

b

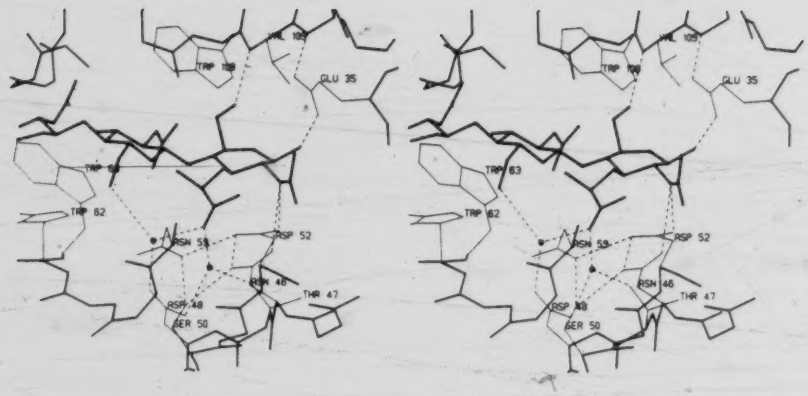


Fig. 6.10 Hydrogen bonding interactions between protein atoms and sugar residues within sites B (a), C (b), and D (c). Lysozyme side chains are shown as thin lines, main chain and sugar as thicker lines. Hydrogen bonds are shown as dashed lines. Water molecules are depicted as black spheres. The hydrogen bond lengths are given in Table 6.5.

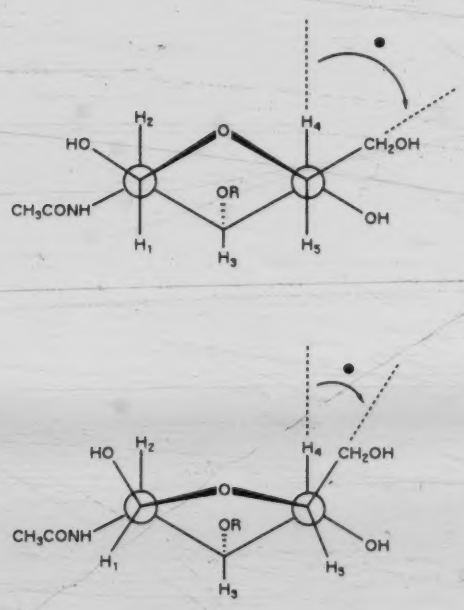


Fig. 6.11 A schematic representation of a NAM sugar residue in the full chair (top) and sofa (bottom) configuration. The change in position of the -CH_2OH group on C5 is from an equatorial to a quasi-axial orientation.

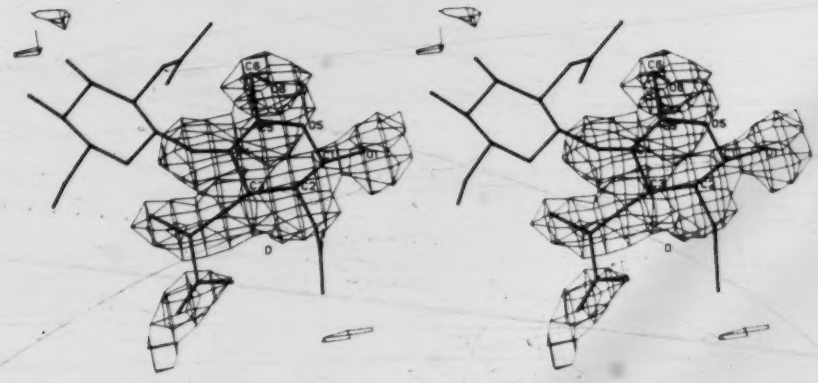


Fig. 6.12 The $|F_o| - |F_c|$, phases α_c, difference electron density resulting after five cycles of restrained least squares refinement on an initial model with the sugar in subsite D removed. The final refined conformation of the subsite D NAM residue we report in this paper is shown in dark lines. The map is contoured at +0.15 e$^-$Å^3.

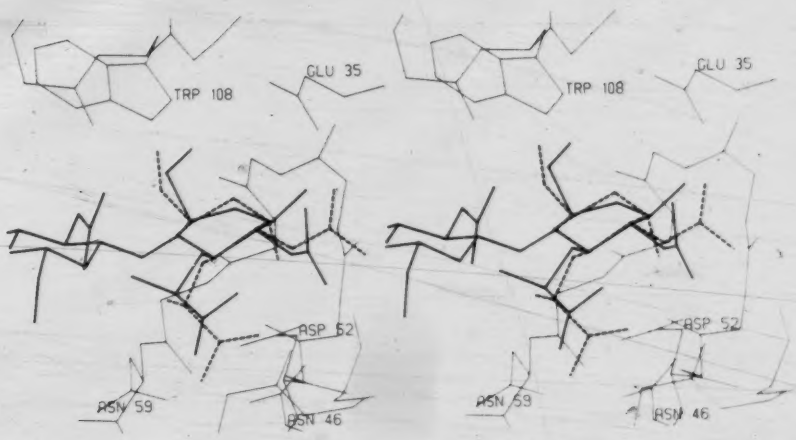
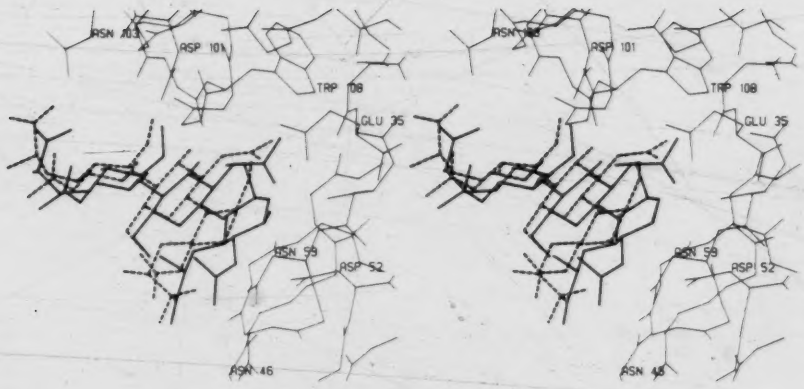
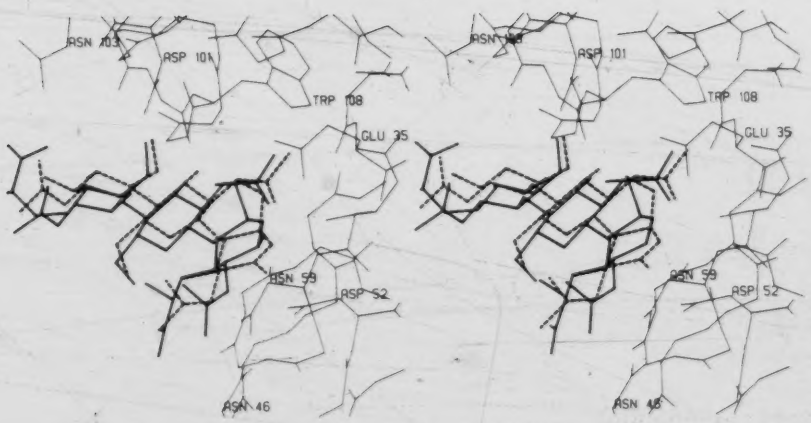


Fig. 6.13 An overlap of the conformation of the N-acetylmuramic residue as bound to subsite D of HEWL (sugar is in solid thick lines; protein is in thin solid lines) with the conformation of 2-acetamido-2-deoxy-α,D-muramic acid determined from crystallographic studies shown in thick dashed lines (Knox & Murthy, 1974).



a



b

Fig. 6.14 A comparison of the relative position and conformation of the NAM-NAG-NAM trisaccharide as determined in this work (thick solid lines) to that of (a) the 2.5 Å resolution, unrefined, structure of the lysozyme bound NAM-NAG-NAM (Kelly *et al.*, 1979) as shown in thick, dashed lines and (b) the 2.5 Å resolution, unrefined structure of the lysozyme bound lactone derivative NAG_3NAL (Ford *et al.*, 1974) also shown in thick dashed lines. In both parts sections of the HEWL molecule are shown relative to the bound saccharide positions.

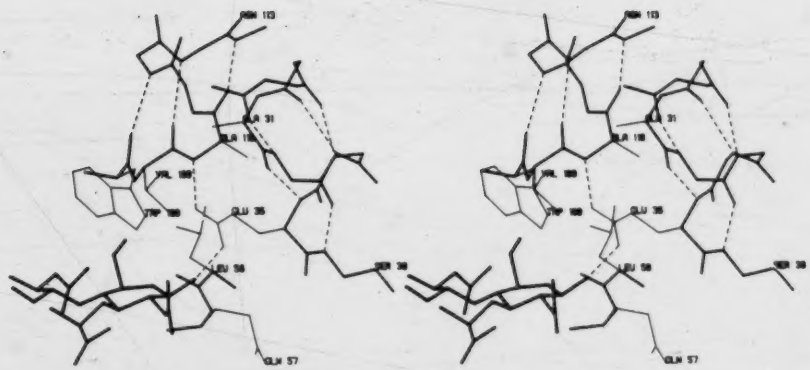


Fig. 6.15 The environment of the catalytic residue Glu35. Main chain and sugar groups are in thick, solid lines, side chains are in thin, solid lines. Hydrogen bonds are shown as thin, dashed lines. Only the lysozyme side chains, critical to the present discussion have been represented so as not to clutter the diagram.

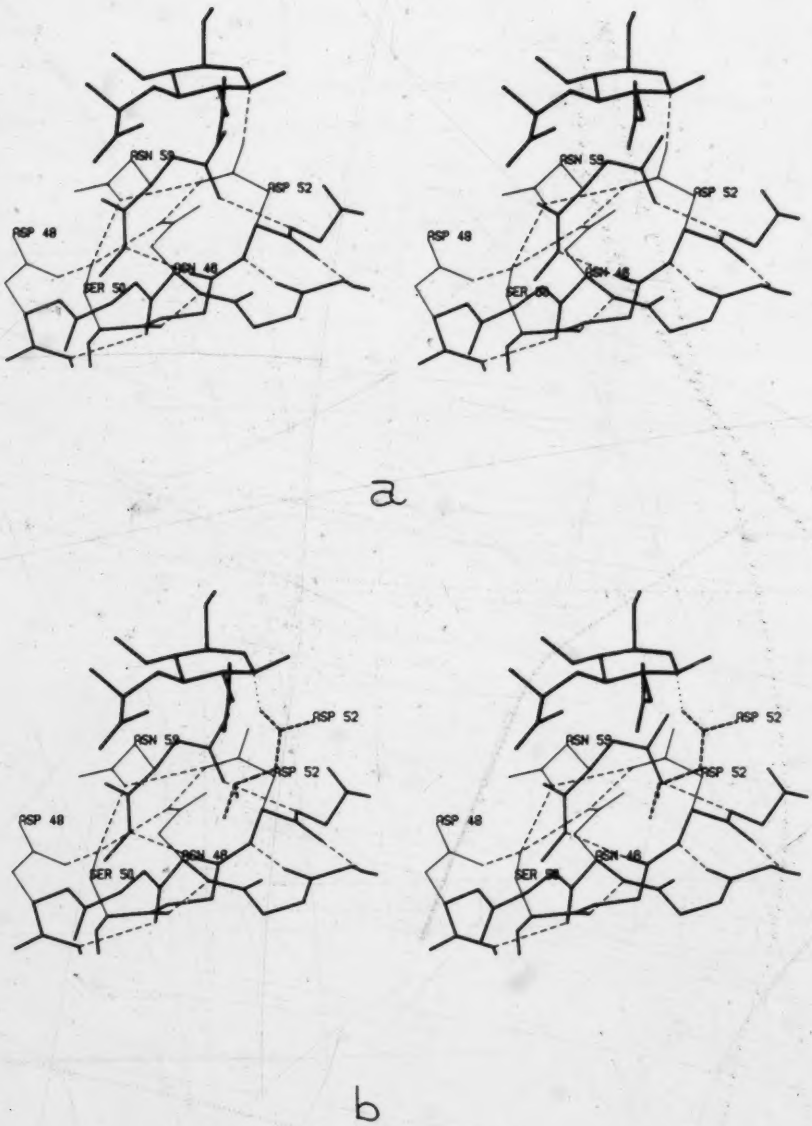


Fig. 6.16 The environment of the catalytic residue Asp52. Main chain and sugar groups are shown as thick, solid lines, side chains as thin, solid lines and hydrogen bonds as thin, dashed lines. Only those lysozyme side-chains that are in the region of Asp52 have been included to ensure clarity. (a) shows the closest approach of the Asp52 carboxylate to the C1 anomeric carbon of the site D sugar which still retains the surrounding, conserved hydrogen bonding networks. (b) shows in thick dashed lines the position Asp52 would have to occupy in order to provide the most favourable stereoelectronic configuration for the promotion of nucleophilic attack on the C1 carbon of the site D sugar. It is obvious from this figure that such a position for Asp52 would require disruption of the hydrogen-bonded network to Asn46 and Asn59 as well as disrupting the surrounding three-stranded β-sheet within which Asp52 is anchored.

CHAPTER 7

DISCUSSION

Several biochemical and biophysical studies pertaining to the work done in the main body of the thesis have been reported subsequent to the publication of these papers. These points along with proposals for possible future directions are presented in the following discussion.

Glycoprotein D of HSV

The validity of targeting glycoprotein D (gD) in the design of a potential vaccine against Herpes simplex virus (HSV) has been reinforced by several recent studies. Evidence that gD plays a role in the early stages of the infection process initially came from the work of Fuller and Spear (1985, 1987) which showed that potent virus-neutralizing monoclonal antibodies against gD were unable to block the attachment of HSV virions to host cells but could inhibit the penetration of the HSV nucleocapsid into the cells. Recently, Johnson and Ligas (1988) have furthered this concept by showing that mutant HSV virions which were unable to express gD were able to attach to cells but could not penetrate into the host cell cytoplasm. Furthermore, transformed mouse and human cell lines constructed to constituitively express HSV gD have been shown to adsorb radiolabelled HSV as effectively as control cell lines but the subsequent step along the infective pathway, i.e. virus induced cell-cell fusion and penetration did not occur. In this way resistance to HSV infection was achieved in these transformed cell lines (Campadelli-Fiume *et al.*, 1988; Johnson & Spear, 1989). It has been proposed on the basis of these studies that gD may be involved in sequestering and interacting with the cell membrane proteins required for fusion.

In addition to the above work there is now added information on the importance of gD as a target for host humoral and cellular immune responses (Martin *et al.*, 1987; Torseth *et al.*, 1987). There are now several documented studies in which both native and recombinant forms of glycoprotein D have been shown to provide protection in various

animal model systems against challenge with herpes simplex virus (Berman *et al.*, 1985; Eisenberg *et al.*, 1985; Lasky *et al.*, 1984; Long *et al.*, 1984; Marsden, 1987; Paoletti *et al.*, 1984; Stanberry *et al.*, 1987).

Since the publication of the work presented in Chapter 2, there have been several additional reports concerning the mapping of antigenic sites on glycoprotein D. Weijer and coworkers, in an approach similar to ours, used a predictive algorithm to locate potential antigenic determinants of gD, synthesized the appropriate peptides, and then used them to test for immunogenicity and neutralization of viral activity (Weijer *et al.*, 1988). This group concluded that peptide sequences 9-21 and 267-281 yielded antisera which reacted strongly with gD in Western blot analysis and showed virus-neutralizing activity *in vitro*. This finding corroborates the result obtained in Chapter 2 for the similar peptides 12-21 and 267-276. The predictive algorithm of Weijer *et al.* did not predict antigenic regions corresponding to our neutralizing peptides 288-297 and 314-323.

Recently Isola *et al.* (1989), using a number of monoclonal antibodies raised against native gD, have shown that the major, continuous antigenic determinants on gD correspond to regions 11-19, 264-279 and 284-301. These results again concur with our findings on the immunogenicity of peptides 12-21, 267-276 and 288-297. In addition, a mutant gD in which proline 273 was replaced by a serine failed to bind to the monoclonal directed against that particular epitope (264-279) in the native structure (Isola *et al.*, 1989). As described in Chapter 2, the epitope containing Pro273 (267-276) was predicted to form a β-turn motif at the molecular surface. Mutation of the proline to a serine could conceivably abolish the geometry required for the turn thus altering the structure of the antigenic determinant.

Clearly the use of synthetic peptides to map the antigenic determinants of structurally undefined proteins has some value. From a primary sequence one can delineate potentially neutralizing, surface regions on a viral protein in the virion bound form. However there are many drawbacks to this methodology.

Most important of these is that antipeptide antibodies can only probe for linear or continuous epitopes on the native protein structure. This is a serious limitation in light of ever increasing evidence from a complement of crystallographic and escape mutant analyses that the majority of immunodominant epitopes on proteins are discontinuous (Mariuzza *et al.*, 1989). These observations in concert with the many conformational isomers of peptides in solution helps to explain the consistently lower neutralizing power of anti-peptide antibodies compared to antibodies directed against the analogous region of the intact parent protein.

In order to obtain the full antigenic picture of glycoprotein D it will be necessary to perform the isolation and sequencing of HSV escape mutants and ideally to carry out an accompanying three dimensional structure determination from X-ray crystallographic or NMR methods. It will be interesting to compare the results from the peptide work to these analyses.

Drug Binding to Calmodulin and Troponin C

One of the major concepts presented in Chapter 3 is the incredibly high degree of structural conservation found among molecules of the helix-loop-helix or EF hand family of proteins. This type of analysis indicates that members within this family should be good candidates for the method of comparative molecular or "homology" modelling (i.e. using the known three-dimensional structure of one member of the family to predict the structure of another member). At the time the drug binding studies on troponin C (TnC) and calmodulin (CaM) were initiated there were no available atomic coordinates for CaM. For this reason a model of CaM was generated using the three dimensional structure of TnC as an initial template (Chapter 4). The eventual report of the full atomic structure of CaM (Babu, 1988) of course provided an immediate means with which to test the accuracy of the model. Least-squares overlap of the two structures yielded r.m.s. differences of .98 for the main chain atoms of the N-terminal domain (residues 1-82) and .79 for the main chain atoms of the C-terminal domain (83-148). The largest differences in the main chain occur,

as postulated in Chapter 4, at the linker regions between helix-loop-helix motifs and in the slight variation of interhelical angle. The side chains of conserved amino acids constituting the calcium binding sites and the hydrophobic core (as outlined in Chapter 3) occupied very similar positions in the computer model and the crystallographic structure. Considering that these are the key areas of interest in the molecule the result is very gratifying. Less conserved residues (often charged amino acids on the molecular surface) superposed less well between the modelled and experimentally determined structures. Overall, the modelling of CaM from TnC is an excellent example of the potential of homology modelling in predicting protein structure.

Both Chapters 4 and 5 propose that the well documented Ca^{2+}-dependent nature of drug binding to TnC and CaM is a result of the exposure of specific hydrophobic residues with which they interact in the Ca^{2+}-free to Ca^{2+}-bound conformational transition. Recently, a number of studies have provided direct evidence for the proposed calcium induced conformational change in TnC and CaM. Grabarek *et al.* (1990) created a TnC mutant in which Gln51 in the B-C linker and Gln85 in the D segment of the central helix were converted to cysteines. These authors reasoned that if the model of Herzberg *et al.* (1986) is correct then creation of a disulphide bridge between these two regions of the protein should inhibit the calcium induced transition. Their results show this to be the case as the oxidized form of the mutant not only had significantly reduced Ca^{2+} affinity in the N-terminal sites I and II but also failed to bind to its molecular target troponin-I, and could not restore Ca^{2+} sensitive ATPase activity to TnC depleted myofibrils.

In a similar study, Fujimora *et al.* (1990) created two TnC mutants in which either Glu57 on helix C or Glu88 on helix D were replaced by a lysine. It was reasoned that such mutations may induce formation of a salt bridge between the two residues which would in turn inhibit the Ca^{2+}-induced conformational change. The reduced Ca^{2+} affinity and decreased levels of calcium activated tension development in reconstituted skinned muscle fibres resulting from these mutants support this proposal. Hoffman & Klevit (1990) have

used 2-D NMR spectroscopy to probe the apo or calcium free form of calmodulin. Preliminary results indicate NOE distances which are consistent with the calcium free model of calmodulin as presented in Chapter 4. Huque and Hogel (1990) have used spin-echo ^{1}H-NMR methodology to follow the oxidation rate of methionine residues of CaM which have been modified by deuterium oxide to the corresponding sulfoxide moiety. The rate of oxidation of several of these residues was shown to be markedly decreased in the apo form of calmodulin. This coincides well with the predicted burial of methionine residues in the model presented in Chapter 4. Finally, Kataoka *et al.* (1989) have used small angle X-ray solution scattering to show that the amphiphilic inhibitory drug mellitin binds to and induces a large conformational change in CaM only in the presence of calcium. Comparison of the distance distribution function, p(r), for isolated and mellitin bound calmodulin indicate that the complex is a more globular structure with a maximum length of approximately 46 Å. This work thus provides supporting evidence for the modelling presented in Chapter 5.

In general there appears to be an increasing amount of experimental support for the calcium dependent interaction of TnC and CaM with various molecular targets as presented in the thesis text. These findings appear to speak well for the use of molecular modelling in probing the nature of molecular interactions. However, more direct validation of these models will require site specific mutagenesis or chemical modification of amino acids proposed to lie at the protein/drug interface and high resolution molecular structures of such complexes as determined by X-ray crystallographic or NMR methodologies.

Understanding the Mechanistic Events Governing the Hydrolytic Activity of Hen Egg White Lysozyme

The X-ray crystallographic analysis at 1.5 Å resolution of the trisaccharide NAM-NAG-NAM bound to hen egg white lysozyme (HEWL) has been informative and gratifying on many levels. Firstly, the structure was able to confirm many earlier suppositions on lysozyme activity made from the elegant modelling studies of Phillips

(1966, 1967). This success again highlights the potential power of molecular modelling in probing molecular interactions. Secondly, the study was a valuable lesson in the amount of additional information a well refined, high resolution structure provides. Over the years many different mechanisms have been proposed to account for lysozyme hydrolytic action (Phillips, 1966; Lowe, 1968; Imoto *et al.*, 1972; Post & Karplus, 1986; Sinnott, 1987). What becomes immediately apparent upon comparison of these various mechanistic proposals is that relatively small differences in molecular conformation and position between enzyme and substrate would either support or negate a given mechanistic event. For this reason, interpretations from unrefined, low resolution, crystallographic structures or from molecular dynamics simulations in which the energy potentials describing atomic motions may not precisely mirror reality should be made with caution.

The structural information gleaned from the 1.5 Å study presented in Chapter 6 not only supports the mechanism of Phillips, but is of sufficient accuracy to provide a more precise underlying molecular basis for the proposed mechanistic events. The formation of a hydrogen bond between the main chain nitrogen of Val109 and the axially placed hydroxymethyl oxygen of the distorted D sugar ring, the formation of a hydrogen bond between the Glu35 carboxylate and the anomeric oxygen on ring D, and the geometrically restricted approach of Asp52 for the anomeric carbon on ring D are examples of such added information.

Future directions in the investigation of the mechanism of lysozyme catalysis is unfortunately severely hampered by the difficult and time consuming task of synthesizing possible substrate and transition state analogs. It would be informative to do the X-ray structure of HEWL complexed to the δ-lactone derivate NAG_3NAL (Ford *et al.*, 1974) refined, and at higher resolution. Other possible substrates might involve nonhydrolyzable analogs; an oligosaccharide in which glycosidic oxygens are replaced by nitrogens is one example.

Site specific mutants of HEWL could also play a role in probing lysozyme action.

Malcolm *et al.* (1987) have shown that replacing Glu35 with Gln or Asp52 with an Asn in HEWL results in total loss of hydrolytic activity. Crystal structures of these mutants with the natural hexasaccharide substrate bound at subsites A through F would be an obvious experiment to aim for. Furthermore, a comparison of the substrates $(NAG)_6$ and $(NAG\text{-}NAM)_3$ bound to HEWL would delineate the differences in the ground state conformation of a NAG or a NAM residue bound at site D. Finally, it would be interesting to look at NAM-NAG-NAM and other substrates bound to lysozyme from other species in order to address the universality of the Phillips mechanism among other members of this enzyme family.

REFERENCES

Babu, Y.S., Bugg, C.E. & Cook, W.J. 1988. *J. Mol. Biol.* **204**, 191-204.

Berman, P.W., Gregory, T., Crase, D. & Lasky, L.A. 1985. *Science* **227**, 1490-1492.

Campadelli-Fiume, G., Arsenakis, F., Farabegoli, F. & Roizman, B. 1988. *J. Virol.* **62**, 159-167.

Eisenberg, R.J., Cerini, C.P., Heliman, C.J., Joseph, A.D., Dietzchold, B., Golub, E., Long, D., Ponce de Leon, M. & Cohen, G.M. 1985. *J. Virol.* **56**, 1014-1017.

Ford, L.O., Johnson, L.N., Machin, P.,A., Phillips, D.C. & Tjian, R. 1974. *J. Mol. Biol.* **88**, 349-371.

Fujimori, K., Sorenson, M., Herzberg, O., Moult, J. & Reinach, F.C. 1990. *Nature* **345**, 182-184.

Fuller, A.O. & Spear, P.G. 1985. *J. Virol.* **55**, 475-482.

Fuller, A.O. & Spear, P.G. 1987. *Proc. Natl. Acad. Sci.* **84**, 5454-5458.

Grabarek, Z., Tan, R-Y., Wang, J., Tao, T. & Gergely, J. 1990. *Nature* **345**, 132135.

Herzberg, O., Moult, J. & James, M.N.G. 1986. *J. Biol. Chem.* **261**, 2638-2644.

Hoffman, R.C., Klevit, R.E. 1990. In *Collected Abstracts of the Seventh International Symposium on Calcium Binding Proteins in Health and Disease*, p. 143.

Huque, E.M. & Hogel, H.J. 1990. In *Collected Abstracts of the Seventh International Symposium on Calcium Binding Proteins in Health and Disease*, p. 142.

Imoto, I., Johnson, L.N., North, A.C.T., Phillips, D.C. & Rupley, J.A. 1972. In *The Enzymes* (Boyer, P., ed.), Vol. VII, 3rd edn., pp. 665-868, Academic Press, New York.

Isola, V., Eisenberg, R.J., Siebert, G.R., Heilman, C.J., Wilcox, W.C. & Cohen, G.H. 1989. *J. Virol.* **63**, 2325-2334.

Johnson, D.C. & Ligas, M.W. 1988. *J. Virol.* **62**, 4605-4612.

Johnson, R.M. & Spear, P.G. 1989. *J. Virol.* **62**, 819-827.

Kataoka, M., Head, J.F., Seaton, B.A. & Engelman, D.M. 1989. *Proc. Natl. Acad. Sci.* **86**, 6944-6948.

Lasky, L.A., Dowbenko, D., Simonsen, C.C. & Berman, P.W. 1984. *Biotechnology* **2**, 527-532.

Long, D., Madara, J., Ponce de Leon, M., Cohen, G.H., Montgomery, P.C. & Eisenberg, R.J. 1984. *Infect. Immun.* **37**, 761-764.

Lowe, G. 1968. *Proc. Roy. Soc. Ser.* **B167**, 431-434.

Mariuzza, R.A., Phillips, S.E.V. & Poljak, R.J. 1987. *Ann. Rev. Biophys. Biophys. Chem.* **16**, 139-159.

Marsden, H.S. 1987. In *Molecular Basis of Virus Disease* (Russel, W.C. & Almond, J.W., eds.), pp. 259-288, Fortieth Symposium of the Society for General Microbiology. Cambridge University, Cambridge.

Martin, S., Moss, B., Berman, P.W., Lasky, L.A. & Rouse, B.T. 1987. *J. Virol.* **61**, 726-734.

Paoletti, E., Lipinskas, B.R., Samsonoff, C., Mercer, S. & Panicali, D. 1984. *Proc. Natl. Acad. Sci.* **81**, 193-197.

Phillips, D.C. 1966. *Scientific American* **215**, 78-90.

Phillips, D.C. 1967. *Proc. Roy. Acad. Sci.* **57**, 484-495.

Post, C.B. & Karplus, M. 1986. *J. Amer. Chem. Soc.* **108**, 1317-1319.

Sinnott, M.L. 1987. In *Enzyme Mechanisms* (Page, M.I. & Williams, A., eds.), pp. 259-296, Royal Soc. Chem., London.

Stanberry, L.R., Bernstein, L.I., Burke, R.L., Pachl, C. & Myers, M.G. 1987. *J. Infect. Dis.* **155**, 914-920.

Torseth, J.W., Cohen, G.H., Eisenberg, R.J., Berman, D.W., Lasky, L.A., Cerini, C.P., Heliman, C.J., Kerwar, S. & Merigan, T.C. 1987. *J. Virol.* **61**, 1532-1539.

Weijer, W.J., Drijfhout, J.W., Geerligs, H.J., Bloemhoff, W., Feijlbrief, M., Bos, C.A., Hoogerhout, P., Kerling, K.E.T., Popken-Boer, T., Slopsema, K., Wilterdink, J.B., Welling, G.W. & Welling-Wester, S. 1988. *J. Virol.* **62**, 501-510.